

The Velocity of the Propagating Wave for Spatially Coupled Systems with Applications to LDPC Codes

Rafah El-Khatib, *Student Member, IEEE* and Nicolas Macris, *Member, IEEE*

Abstract

We consider the dynamics of message passing for spatially coupled codes and, in particular, the set of density evolution equations that tracks the profile of decoding errors along the spatial direction of coupling. It is known that, for suitable boundary conditions and after a transient phase, the error profile exhibits a “solitonic behavior”. Namely, a uniquely-shaped wavelike solution develops, that propagates with constant velocity. Under this assumption we derive an analytical formula for the velocity in the framework of a continuum limit of the spatially coupled system. The general formalism is developed for spatially coupled low-density parity-check codes on general binary memoryless symmetric channels which form the main system of interest in this work. We apply the formula for special channels and illustrate that it matches the direct numerical evaluation of the velocity for a wide range of noise values. A possible application of the velocity formula to the evaluation of finite size scaling law parameters is also discussed. We conduct a similar analysis for general scalar systems and illustrate the findings with applications to compressive sensing and generalized low-density parity-check codes on the binary erasure or binary symmetric channels.

Index Terms

Message passing, density evolution, potential functional, threshold saturation, soliton, wave propagation, compressive sensing.

I. INTRODUCTION

Spatial coupling was first introduced by Felstrom and Zigangirov in the context of low-density parity-check (LDPC) codes [1]. Spatially coupled codes have been shown to be capacity-achieving on binary memoryless symmetric (BMS) channels under belief propagation (BP) decoding. The capacity-achieving property is due to the “threshold saturation” of the BP threshold of the coupled system towards the maximum a-posteriori (MAP) threshold of the uncoupled code ensemble [2], [3]. Spatial coupling has also been applied to several other problems besides channel coding [4], [5], such as lossy source compression [6], [7], compressive sensing [8], [9], [10], random constraint satisfaction problems [11], [12], [13], and a coupled Curie-Weiss (toy) model [14], [15].

R. El-Khatib and Nicolas Macris are with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, e-mails: {rafah.el-khatib,nicolas.macris}@epfl.ch.

Consider a single (uncoupled) LDPC code. To construct the corresponding coupled code of spatial length $L_c + w$, we take $L_c + w$ replicas of the single system and “couple” every w adjacent single systems by means of a uniform window function. At every iteration of the BP algorithm, the variable and check nodes of the coupled graph exchange messages which are described by a set of coupled density evolution (DE) iterative equations. The solution to the DE equations, called the *decoding profile* \mathbf{x} , is a vector of “error distributions” (more precisely, distributions of the BP log-likelihood estimates) along the spatial axis of coupling. More specifically, let the integer $z \in \{-w + 1, \dots, L_c\}$ denote the position along the spatial direction of the graph construction (on which the replicas are spread). Then the z^{th} component of \mathbf{x} , call it x_z , denotes the distribution of the BP log-likelihood estimate at the z^{th} position. In the special case of the binary erasure channel (BEC) this component is reduced to the usual scalar erasure probability $0 \leq x_z \leq 1$ at position z along the spatial axis of coupling.

Spatially coupled codes perform well, and are capacity achieving, due to the threshold saturation phenomenon that is proved for general BMS channels in [2], [3]. More specifically, as long as the channel noise is below the MAP threshold, the decoding profile of a *spatially coupled* code converges to the all- Δ_∞ vector after enough iterations of the BP algorithm, where Δ_∞ is the Dirac mass at infinite log-likelihood (i.e. perfect knowledge of the bits). In the special case of the BEC, the all- Δ_∞ vector corresponds to a vector of scalar erasure probabilities driven to zero by DE iterations. On the other hand, the probability distribution of the log-likelihoods of bits of the corresponding *uncoupled* code only converge to Δ_∞ when the channel noise is below the BP threshold (which is lower than the MAP threshold).

The threshold saturation phenomenon is made possible due to “seeding” at the boundaries of the spatially coupled code. Seeding means that we fix the bits at the boundaries so that the probability distributions of their log-likelihoods are Δ_∞ . This facilitates BP decoding near the boundaries, and this effect is propagated along the rest of the coupled chain. The minimum size of the seed that guarantees the propagation of the decoding effect is of the same order as the size w of the coupling window; however, an exact determination of the minimum possible such size is an still an interesting open question.

When the channel noise is between the BP and the MAP thresholds, and the underlying uncoupled ensemble has a unique non-trivial stable BP fixed point that blocks decoding, an interesting phenomenon that has been empirically observed is the appearance of a *solitonic decoding wave* after a certain number of transient iterations of the BP algorithm. This *soliton* is characterized by a *fixed shape* that seems independent of the initial condition and has a *constant traveling velocity* that we henceforth call v . This phenomenology is discussed in more details in Section II-C. Figures 2 and 3 in Section II-C show an example of the transient phase and of the soliton in the case of spatially coupled codes on the BEC. The main goal of this work is to derive a formula for the velocity of the soliton for *general BMS channels*.

The decoding wave has recently been studied in the context of coding when transmission takes place over the BEC. In [16], it is proved that the solitonic wave solution exists and bounds on the velocity of the soliton are derived. However, the independence of the unique shape of the wave from the initial conditions remains an open question. In [17], more complex coupled systems are studied, where it is possible to have more than one non-trivial stable BP fixed point, and there again some bounds on the velocity of the soliton are provided. The solitonic

behavior has also been studied for the coupled Curie-Weiss toy model [14] and in [15] a formula for the velocity, as well as an approximation, are derived and tested numerically.

In the first part of this work, we derive a general formula for the velocity of the wave in the asymptotic limit $L_c \gg w \gg 1$ in the context of coding when transmission takes place over *general* BMS channels (see Equ. (15)). This limit enables us to formulate the problem in a “continuum limit” which makes the derivations quite tractable. We show, with the use of numerical simulations, that this continuum limit yields good approximations for the velocity of the original discrete system. For simplicity, we limit ourselves to the case where the underlying uncoupled LDPC code has only *one non-trivial stable BP fixed point*.

Our derivation rests on the assumption that the soliton indeed appears. More precisely, we assume that after an initial transient phase, the decoding profile develops a unique shape, independent of the initial condition, and travels with a constant velocity v . This assumption can be strictly true only in an asymptotic limit of a very large chain length and a large iteration number (or time). It is an interesting open problem to make this space-time asymptotic limit precise and rigorously prove that the soliton appears and is independent of the initial condition. We conjecture that our velocity formula is exact in such a limit.

The formula for the velocity of the wave greatly simplifies when we consider transmission over the BEC, because the decoding profile reduces to a scalar vector of erasure probabilities. For transmission over general BMS channels, we also simplify the analysis by applying the Gaussian approximation [18], [19]. This consists of approximating the DE densities and the channel by suitable “symmetric” Gaussian densities. Since the mean m and the variance σ^2 of these symmetric Gaussian densities are related by $\sigma^2 = 2m$, the analysis then reduces to that of a one-dimensional scalar system, whose technical difficulty is similar to that of the case of transmission over the BEC. We thus obtain a more tractable velocity formula and compare the numerical predictions of these velocity formulas with the empirical value of the velocity for finite values of L_c and w . Good agreement is found, on practically the whole range of values within $[\epsilon_{BP}, \epsilon_{MAP}]$, even for small values of the window size w .

It is of theoretical as well as practical interest to have a hold on the analytical expression of the velocity of the wave. The velocity is also related to other fundamental quantities that describe a coding system, such as the finite-size scaling law that predicts the error probability of finite-length spatially coupled codes. In [20], the scaling law for a finite-length spatially coupled (ℓ, r, L_c) code, when transmission takes place over the BEC, is derived. Involved in this scaling law are parameters that can be estimated using the value of the velocity of the decoding wave. Using values of the velocity computed in our work, we provide reasonably good estimates of these parameters.

In the second part of this work, we consider general spatially coupled scalar bipartite systems (that are not restricted to coding) governed by a general message passing algorithm. In this setting, the system is scalar (one-dimensional) since the messages exchanged between the nodes are scalar. Due to seeding at the boundary, the “profile” (we no longer call it the “decoding profile”) exhibits the same phenomenology as in coding. Namely, a solitonic behavior appears after a short transient phase. We derive a formula for the velocity of the soliton for such systems and illustrate it on two applications: compressive sensing and generalized LDPC (GLDPC) codes.

The derivations of the velocity formulas in both parts of the work use the same tools and assumptions. We combine the use of the “*potential functional*” introduced and used in a series of works [3], [11], [14], [21], [22],

as well as the continuum limit $L_c \gg w \gg 1$ which makes the derivations analytically tractable. The potential is a “variational formulation” of the message passing algorithm on coding systems. It is a functional whose stationary points are the fixed points of the density evolution equations described by this algorithm, and has been used to prove threshold saturation in [3] for general BMS channels. A significant part of the formalism in [3] is used in the present work. We also note that potential formulations have been used to characterize the fixed point(s) of general scalar systems at the MAP threshold using displacement convexity in [23], [24]. An extension of the ideas in these works could shed some light on the question of the independence of the soliton’s shape from the initial conditions.

Section II introduces a few preliminary notions that we will need and reviews the phenomenology of the solitonic wave. In Section III, we formulate the continuum limit and state our main formula for the velocity of the soliton on general BMS channels; the derivation is presented in Section IV. Comparisons with numerical experiments are presented in Section V. These concern transmission over the BEC as well as general BMS channels in the so-called Gaussian approximation. We also discuss a possible application of our formula to scaling laws for finite-size ensembles in Section VI. The case of general scalar spatially coupled systems is treated in Section VII, and illustrated for the examples of generalized LDPC codes (on the BEC or BSC channels) and compressive sensing. We present concluding remarks and propose further directions in Section VIII.

A summary of this work has appeared in [25], [26].

II. PRELIMINARIES

We consider (almost) the same setting as in [3] and adopt most of the notation introduced in that work. For more information about the formalism in these preliminaries, one can consult [27].

We denote by $M(\bar{\mathbb{R}})$ the space of probability measures \mathbf{x} on the extended real numbers $\alpha \in \bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. Here $\alpha \in \bar{\mathbb{R}}$ should be interpreted as a “log-likelihood variable”. We call the measure \mathbf{x} *symmetric* if $\int_E d\mathbf{x}(\alpha) = \int_{-E} e^{-\alpha} d\mathbf{x}(\alpha)$ for all measurable sets $E \subset \bar{\mathbb{R}}$.

We define an *entropy functional* $H : \mathcal{M} \rightarrow \mathbb{R}$ that maps a finite probability measure from $M(\bar{\mathbb{R}})$ to a real number. It is defined as

$$H(\mathbf{x}) = \int d\mathbf{x}(\alpha) \log_2(1 + e^{-\alpha}). \quad (1)$$

Note that this is a linear functional. Linearity is used in an important way to compute the entropy of convex combinations of measures (which also yield probability measures). But we will also compute the “entropy” associated to differences of measures by setting $H(\mathbf{x}_1 - \mathbf{x}_2) \equiv H(\mathbf{x}_1) - H(\mathbf{x}_2)$. In other words, the entropy functional is extended in an obvious way to the space of signed measures.

In the remainder of this work, we will use the Dirac masses $\Delta_0(\alpha)$ and $\Delta_\infty(\alpha)$ at zero and infinite log-likelihood, that have entropies $H(\Delta_0) = 1$ and $H(\Delta_\infty) = 0$, respectively.

We will also use the standard variable-node and check-node convolution operators \oplus and \boxtimes for log-likelihood ratio message distributions involved in DE equations [27]. For $\mathbf{x}_1, \mathbf{x}_2 \in M(\bar{\mathbb{R}})$, the usual convolution $\mathbf{x}_1 \oplus \mathbf{x}_2$ is the density of

$$\alpha = \alpha_1 + \alpha_2,$$

and $\mathbf{x}_1 \boxtimes \mathbf{x}_2$ is the density of

$$\alpha = 2 \tanh^{-1} \left(\tanh \frac{\alpha_1}{2} \tanh \frac{\alpha_2}{2} \right).$$

More formally, for any measurable set $E \in \mathbb{R}$, the operators are defined by

$$\begin{cases} (\mathbf{x}_1 \otimes \mathbf{x}_2)(E) &= \int d\mathbf{x}_2(\alpha) \mathbf{x}_1(E - \alpha), \\ (\mathbf{x}_1 \boxtimes \mathbf{x}_2)(E) &= \int d\mathbf{x}_2(\alpha) \mathbf{x}_1 \left(2 \tanh^{-1} \left(\frac{\tanh(E/2)}{\tanh(\alpha/2)} \right) \right). \end{cases} \quad (2)$$

We note that the identities of the \otimes and \boxtimes operators are Δ_∞ and Δ_0 , and their annihilators are Δ_0 and Δ_∞ . More explicitly,

$$\begin{cases} \Delta_\infty \boxtimes \mathbf{x} = \mathbf{x}, & \Delta_0 \boxtimes \mathbf{x} = \Delta_0, \\ \Delta_0 \otimes \mathbf{x} = \mathbf{x}, & \Delta_\infty \otimes \mathbf{x} = \Delta_\infty. \end{cases} \quad (3)$$

Each operation, taken separately, is associative, commutative, and linear. However, when they are taken together, there is no distributive law; also, they don't associate in the sense that $\mathbf{x}_1 \otimes (\mathbf{x}_2 \boxtimes \mathbf{x}_3) \neq (\mathbf{x}_1 \otimes \mathbf{x}_2) \boxtimes \mathbf{x}_3$ and $\mathbf{x}_1 \boxtimes (\mathbf{x}_2 \otimes \mathbf{x}_3) \neq (\mathbf{x}_1 \boxtimes \mathbf{x}_2) \otimes \mathbf{x}_3$. We will also use the so-called *duality rules*

$$\begin{cases} H(\mathbf{x} \otimes \mathbf{y}) + H(\mathbf{x} \boxtimes \mathbf{y}) = H(\mathbf{x}) + H(\mathbf{y}), \\ H(\mathbf{x} \otimes \mathbf{a}) + H(\mathbf{x} \boxtimes \mathbf{a}) = H(\mathbf{a}), \\ H(\mathbf{a} \otimes \mathbf{b}) + H(\mathbf{a} \boxtimes \mathbf{b}) = 0, \end{cases} \quad (4)$$

where $\mathbf{x}, \mathbf{y} \in M(\bar{\mathbb{R}})$ and \mathbf{a}, \mathbf{b} are differences of probability measures $\mathbf{a} = \mathbf{x}_1 - \mathbf{x}_2$, $\mathbf{b} = \mathbf{x}_3 - \mathbf{x}_4$, $\mathbf{x}_i \in M(\bar{\mathbb{R}})$, $i = 1, 2, 3, 4$.

A. Single System

Consider an (uncoupled) LDPC(λ, ρ) code ensemble and transmission over the BMS channel. Here, $\lambda(y) = \sum_l \lambda_l y^{l-1}$ and $\rho(y) = \sum_r \rho_r y^{r-1}$ are the usual edge-perspective variable-node and check-node degree distributions, respectively. The node-perspective degree distributions L and R are defined by $L'(y) = L'(1)\lambda(y)$ and $R'(y) = R'(1)\rho(y)$, respectively. Moreover, consider communication over a family of BMS channels whose distribution $c_h(\alpha)$ in the log-likelihood domain is parametrized by the channel entropy¹ $H(c_h) = h$.

Let $\tilde{\mathbf{x}}^{(t)}$ denote the variable-node output distribution of the BP algorithm at iteration $t \in \mathbb{N}$. We can track the average behavior of the BP decoder by means of the DE iterative equations that are written as a recursion in terms of the variable-node output distribution as follows

$$\tilde{\mathbf{x}}^{(t+1)} = c_h \otimes \lambda^{\otimes}(\rho^{\boxtimes}(\tilde{\mathbf{x}}^{(t)})), \quad (5)$$

with initial condition $\tilde{\mathbf{x}}^{(0)} = \Delta_0$ (equivalently, we can take the perhaps more natural initial condition $\tilde{\mathbf{x}}^{(0)} = c_h$).

There are two thresholds of interest for us. The first one is the algorithmic threshold; it is defined for a family of BMS channels whose channel distributions $c_h(\alpha) : \mathbb{R} \rightarrow M(\bar{\mathbb{R}})$ are ordered by degradation and parametrized by their entropy $H(c_h) = h$. It is also called the BP threshold of the family and is defined as

$$h_{\text{BP}} = \{h \in [0, 1] : \tilde{\mathbf{x}} = c_h \otimes \lambda^{\otimes}(\rho^{\boxtimes}(\tilde{\mathbf{x}})) \implies \tilde{\mathbf{x}} = \Delta_\infty\}.$$

¹In the literature, this quantity is often denoted by $c(h)$.

The second threshold corresponds to optimal (MAP) decoding

$$\mathbf{h}_{\text{MAP}} = \{\mathbf{h} \in [0, 1] : \liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[H(X^n | Y^n(\mathbf{h}))] > 0\},$$

where $H(X^n | Y^n(\mathbf{h}))$ is the conditional Shannon entropy of the input given by the channel observations, and \mathbb{E} is the expectation over the code ensemble.

The potential functional $W_s(\mathbf{x})$, $\mathbf{x} \in \mathcal{M}(\bar{\mathbb{R}})$, of the “single” or uncoupled system is

$$W_s(\mathbf{x}) = \frac{1}{R'(1)} H(R^{\boxtimes}(\mathbf{x})) + H(\rho^{\boxtimes}(\mathbf{x})) - H(\mathbf{x} \boxtimes \rho^{\boxtimes}(\mathbf{x})) - \frac{1}{L'(1)} H(\mathbf{c} \otimes L^{\otimes}(\rho^{\boxtimes}(\mathbf{x}))). \quad (6)$$

The fixed point form of the DE equation (5) is obtained by setting to zero the functional derivative of $W_s(\mathbf{x}; \mathbf{c})$ with respect to \mathbf{x} . In other words, $\mathbf{x} = \mathbf{c}_h \otimes \lambda^{\otimes}(\rho^{\boxtimes}(\mathbf{x}))$ is equivalent to

$$\lim_{\gamma \rightarrow 0} \frac{1}{\gamma} (W_s(\mathbf{x}) + \gamma \eta) - W_s(\mathbf{x}) = 0, \quad (7)$$

where η is a difference of two probability measures (see [3] for the proof of this statement). The BP and MAP thresholds, \mathbf{h}_{BP} and \mathbf{h}_{MAP} , respectively, can be obtained from the analysis of the stationary points of the potential function. See [3], [28] for more details and a rigorous discussion of this issue.

Remark about notation. In the remainder of this work, most of the time, we omit the subscript \mathbf{h} from \mathbf{c}_h and the argument α from $\mathbf{x}(\alpha)$. This is because we will need a subscript (resp. an argument) z that represents the position along the chain in the discrete (resp. continuous) case.

B. Spatially Coupled System

For standard LDPC codes, the BP threshold \mathbf{h}_{BP} is, in general, lower than the MAP threshold \mathbf{h}_{MAP} . The definitions of the BP and MAP thresholds above extend to the spatially coupled setting. Spatial coupling exhibits two attractive properties. First, *the MAP threshold is conserved* under coupling in the limit $L_c \rightarrow +\infty$ and for all w . The proof of this statement is found in [28] (see also [29], [30]). Second, *the BP threshold of the coupled system saturates to its MAP threshold* as proved in [2], [3]. The main consequence of threshold saturation is that one can decode perfectly up to the \mathbf{h}_{MAP} .

Let us now describe the density evolution and potential functional formalism for the spatially coupled code ensemble. Consider $L_c + w$ “replicas” of the single system described in Section II-A, placed on the spatial coordinates $z \in \{-w + 1, \dots, L_c\}$. The system at position z is coupled to w neighboring systems by means of a coupling window. For simplicity, we consider a uniform coupling window. We denote by $\tilde{\mathbf{x}}_z^{(t)}$ the variable-node output distribution at position $z \in \{-w + 1, \dots, L_c\}$ on the spatial axis and at time $t \in \mathbb{N}$. The DE equation of the coupled system takes the form

$$\tilde{\mathbf{x}}_z^{(t+1)} = \mathbf{c}_z \otimes \lambda^{\otimes} \left(\frac{1}{w} \sum_{i=0}^{w-1} \rho^{\boxtimes} \left(\frac{1}{w} \sum_{j=0}^{w-1} \tilde{\mathbf{x}}_{z+i-j}^{(t)} \right) \right). \quad (8)$$

In this equation, $\mathbf{c}_z = \mathbf{c}$, for $z \in \{1, \dots, L_c\}$ and $\mathbf{c}_z = \Delta_\infty$ for $z \in \{-w + 1, \dots, 0\}$. Furthermore, we fix the left boundary to $\mathbf{x}_z^{(t)} = \Delta_\infty$ for $z \in \{-w + 1, \dots, 0\}$, for all $t \in \mathbb{N}$. These conditions express perfect information

at the left boundary which is what enables seeding and the decoding wave propagation along the chain of coupled codes. The initial condition (8) is $\mathbf{x}_z^{(0)} = \Delta_0$ for $z \in \{1, \dots, L_c + w - 1\}$.

It will be convenient to work with a smoothed version of the profile $\tilde{\mathbf{x}}_z^{(t)}$, namely $\mathbf{x}_z^{(t)} = \frac{1}{w} \sum_{i=0}^{w-1} \tilde{\mathbf{x}}_{z-i}^{(t)}$, which is the check-node input distribution. Then, using this change of variables, (8) can be rewritten as

$$\mathbf{x}_z^{(t+1)} = \frac{1}{w} \sum_{i=0}^{w-1} \mathbf{c}_{z-i} \otimes \lambda^{\otimes} \left(\frac{1}{w} \sum_{j=0}^{w-1} \rho^{\boxtimes}(\mathbf{x}_{z-i+j}^{(t)}) \right). \quad (9)$$

Just as in the single system case, this DE equation can be expressed as the stationarity condition of a potential functional (see [3])

$$W(\underline{\mathbf{x}}) = \sum_{z=-w+1}^L \left\{ \frac{1}{R'(1)} H(R^{\boxtimes}(\mathbf{x}_z)) + H(\rho^{\boxtimes}(\mathbf{x}_z)) - H(\mathbf{x}_z \boxtimes \rho^{\boxtimes}(\mathbf{x}_z)) - \frac{1}{L'(1)} H\left(\mathbf{c}_z \otimes L^{\otimes} \left(\frac{1}{w} \sum_{i=0}^{w-1} \rho^{\boxtimes}(\mathbf{x}_{z+i}) \right) \right) \right\}. \quad (10)$$

where $\underline{\mathbf{x}} = (\mathbf{x}_{-w+1}, \dots, \mathbf{x}_{L+w-1})$. The fixed point form of (9) is equivalent to $\lim_{\gamma \rightarrow 0} \gamma^{-1}(W(\underline{\mathbf{x}} + \gamma \underline{\eta}) - W(\underline{\mathbf{x}})) = 0$ for $\underline{\eta} = (\eta_{-w+1}, \dots, \eta_{L+w-1})$ where η_i are differences of probability measures.

C. Phenomenological observations

Our derivation is far from rigorous and is based on an assumption derived from a phenomenological picture observed from simulations. We summarize the main observations in this paragraph for the case of transmission over the BEC channel. This channel also gives us the opportunity to illustrate the formalism outlined in Sections II-A and II-B in a concrete case.

The BEC has channel distribution $\mathbf{c}_\epsilon(\alpha) = \epsilon \Delta_0 + (1 - \epsilon) \Delta_\infty$, where ϵ is the erasure probability, and $H(\mathbf{c}_\epsilon) = \epsilon$ (hence $\mathbf{h} = \epsilon$). The density of the BP estimates of log-likelihood variables can be parametrized as $\mathbf{x}^{(t)}(\alpha) = x^{(t)} \Delta_0(\alpha) + (1 - x^{(t)}) \Delta_\infty(\alpha)$, where $x^{(t)} \in [0, 1]$ is interpreted as the erasure probability at iteration $t \in \mathbb{N}$. The DE equation becomes a one-dimensional iterative map

$$x^{(t+1)} = \epsilon \lambda(1 - \rho(1 - x^{(t)})) \quad (11)$$

over scalars in $[0, 1]$. These iterations are always initialized with $x^{(0)} = 1$ or, equivalently, $x^{(0)} = \epsilon$. The corresponding fixed point equation is the stationarity condition for the potential function

$$W_{\text{BEC}}(x) = \frac{1}{R'(1)}(1 - R(1 - x)) - x\rho(1 - x) - \frac{\epsilon}{L'(1)}L(1 - \rho(1 - x)). \quad (12)$$

Note that the potential function is defined up to a constant which is set here so that $W_{\text{BEC}}(0) = 0$. Figure 1 illustrates the potential function for a (3, 6)-regular Gallager ensemble, for several values of ϵ . For $\epsilon < 0.4294$, the potential function (12) is strictly increasing, and equivalently the DE iterations are driven to the unique minimum at $x = 0$. At $\epsilon_{\text{BP}} = 0.4294$ a horizontal inflexion point appears and a second non-trivial local minimum x_{BP} appears; this minimum corresponds to the non-trivial fixed point reached by DE iterations. It is known that the MAP threshold is equal to the erasure probability where the non-trivial local minimum is at the same height as the trivial one and that decoding becomes impossible once the non-trivial minimum becomes a global minimum. For this example, this happens when $\epsilon_{\text{MAP}} = 0.4881$. Figure 1 also shows the *energy gap* that is defined for $\epsilon_{\text{BP}} \leq \epsilon \leq \epsilon_{\text{MAP}}$ as $\Delta E = W_{\text{BEC}}(x_{\text{BP}}) - W_{\text{BEC}}(0)$. At the MAP threshold, we have $\Delta E = 0$.

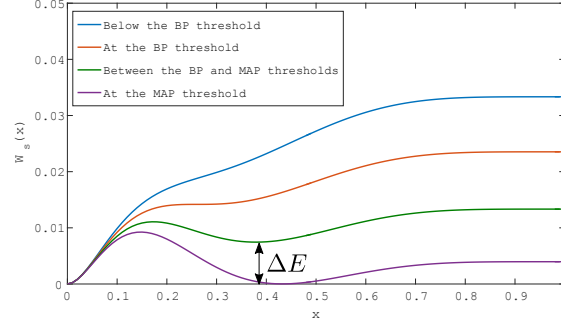


Fig. 1. We plot the potential function of the $(3,6)$ regular LDPC when transmission takes place over the $\text{BEC}(\epsilon)$ for several values of the erasure probability ϵ . Note that $x = 0$ is always a trivial stationary point. For $\epsilon < 0.4294 = \epsilon_{\text{BP}}$ the potential function is strictly increasing and $x = 0$ is the only minimum. At $\epsilon_{\text{BP}} = 0.4294$ a horizontal inflexion point develops and a second non-trivial minimum exists for larger ϵ . At $\epsilon_{\text{MAP}} = 0.4881$ the trivial and non-trivial minimum are at the same height and the energy gap ΔE vanishes.

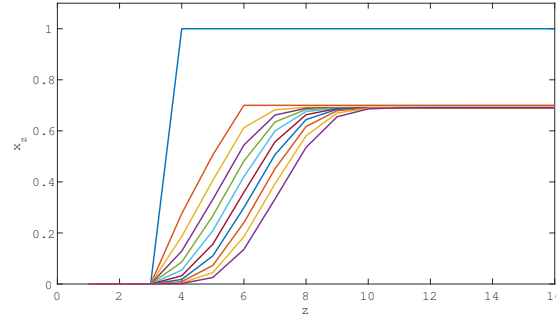


Fig. 2. We consider the spatially coupled $\text{LDPC}(\lambda(x), \rho(x))$ ensemble, with $\lambda(x) = 0.3x^3 + 0.4x^5 + 0.3x^6$ and $\rho(x) = x^5$, with transmission over the $\text{BEC}(0.7)$. The parameters of coupling are $L_c + w = 16$ and $w = 3$. We plot the decoding profile during the first 10 iterations, where the profile is initialized to 0 when $z \leq 3$ and to 1 elsewhere. We observe that the segment initialized to 1 decreases quickly and converges to the BP threshold $x_{\text{BP}} = 0.6907$. We observe that the transient phase is only about 5 iterations here, before the decoding profile converges to a fixed shape.

Let us now describe the phenomenology of the soliton (decoding wave) for spatially coupled codes. Our discussion is limited to the case where the underlying code ensemble has a single non-trivial DE fixed point (equivalently, the potential function has a single non-trivial local minimum). One can show that this is always the case for regular code ensembles. For irregular degree distributions the situation may be more complicated with many non-trivial fixed points that appear. For transmission over the BEC, equation (9) reads

$$x_z^{(t+1)} = \frac{1}{w} \sum_{i=0}^{w-1} \epsilon_{z-i} \lambda \left(\frac{1}{w} \sum_{j=0}^{w-1} (1 - \rho(1 - x_{z-i+j}^{(t)})) \right). \quad (13)$$

Here, $\epsilon_z = 0$ for $z \in \{-w+1, \dots, 0\}$ and $\epsilon_z = \epsilon$ for $z \in \{1, \dots, L_c\}$. Furthermore, we fix the left boundary to $x_z^{(t)} = 0$ for $z \in \{-w+1, \dots, 0\}$, for all $t \in \mathbb{N}$. These are the “perfect seeding” conditions which enable the initiation of decoding. The initial condition for the iterations is $x_z^{(0)} = 1$ (or ϵ) for $z \in \{1, \dots, L_c + w - 1\}$.

The evolution of the decoding wave can be decomposed into two phases: a *transient* and a *stationary* phase. In the

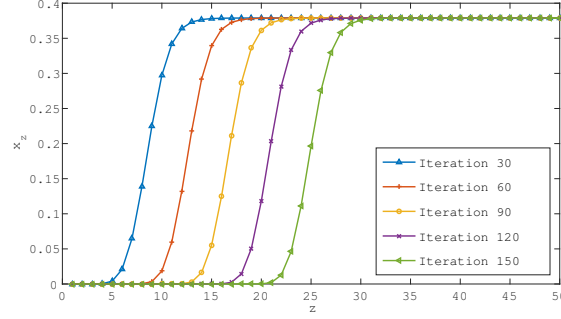


Fig. 3. We consider the $(3,6)$ -regular LDPC spatially coupled code with $L = 50$, $w = 3$ on the BEC(0.46). We plot the error probability along the spatial dimension and observe the “decoding wave”. This “soliton” is plotted every 30 iterations until iteration 150 and is seen to make a quick transition from zero error probability to the BP-value $x_{\text{BP}} = 0.3798$ of the error probability. The optimal (MAP) noise threshold is $\epsilon_{\text{MAP}} = 0.4881$.

transient phase, we observe a profile of erasure probabilities x_z changing shape. The segment initialized to $x_z^{(0)} = 1$ quickly drops to $x_z \approx x_{\text{BP}}$ where it remains stuck on the far right for large values of z . The seeding region, on the other hand, starts progressing towards the right-hand side and, after a few iterations, a fixed profile shape develops. This transient phase is illustrated in Figure 2 for an irregular code. Overall, it only lasts for a few iterations (of the order of 5 iterations in this example). After this transient phase is over, one observes a stationary phase with a *solitonic behavior*, as depicted in Figure 3. The profile of erasure probabilities has a stationary shape with a front at position z_{front} that moves, at a constant speed, towards the right. The soliton is relatively well-localized within approximately $2w$ positions and quickly approaches $x_z \rightarrow 0$ for $z < z_{\text{front}}$ and $x_z \rightarrow x_{\text{BP}}$ for $z > z_{\text{front}}$. The stationary phase and its soliton are depicted in Figure 3 for a finite spatially coupled $(3,6)$ -regular ensemble with chain length $L_c = 50$ and $w = 3$ for $\epsilon = 0.46$. In this figure, we plot the decoding profile every 30 iterations starting at the 30th iteration (the leftmost curve) and until the 150th iteration (the rightmost curve). The kink increases sharply from $x_z = 0$ to $x_z = x_{\text{BP}} = 0.3789$ over a width of the order of $2w = 6$.

III. CONTINUUM LIMIT AND MAIN RESULT

A. Continuum Limit

We now consider the coupled system in the *continuum limit*, in which the length of the coupling chain L_c is first taken very large $L_c \rightarrow +\infty$, and then the window size is taken very large $w \rightarrow +\infty$. The continuum limit has already been considered for the special case of the BEC in [16], [23], [24]. We slightly abuse notation by keeping the same symbols for the profile, the spatial position, and the channel distribution in the continuum limit. Thus, we denote by $\mathbf{x}(\cdot, \cdot)$ the *continuous* profile of distributions and set $\mathbf{x}(\frac{z}{w}, t) \equiv \mathbf{x}_z^{(t)}$. We then replace $\frac{z}{w} \rightarrow z$ so that the new z is the *continuous* variable on the spatial axis, $z \in \mathbb{R}$.

In view of the discussion of the phenomenology in Section II-C, we consider the class of profiles satisfying the “natural boundary conditions” $\mathbf{x}(z, t) \rightarrow \Delta_\infty$ when $z \rightarrow -\infty$ for all $t \in \mathbb{R}$, $\mathbf{x}(z, t) \rightarrow \mathbf{x}_{\text{BP}}$ when $z \rightarrow +\infty$ for all $t \in \mathbb{N}$, where \mathbf{x}_{BP} is the unique non-trivial stable fixed point of DE for the single system Equ. (5).

The BMS channel distribution is now also continuous, and we denote by $c(z)$ the channel distribution at the continuous spatial position $z \in \mathbb{R}$. The DE equation (9) then takes the form

$$\mathbf{x}(z, t+1) = \int_0^1 du \, c(z-u) \otimes \lambda^{\otimes} \left(\int_0^1 ds \, \rho^{\boxtimes}(\mathbf{x}(z-u+s, t)) \right). \quad (14)$$

The initial condition at $t=0$ is given by a profile $\mathbf{x}(z, 0)$ that interpolates between the two limiting values of the boundary condition, namely $\mathbf{x}(z, 0) \rightarrow \Delta_\infty$ when $z \rightarrow -\infty$ and $\mathbf{x}(z, 0) \rightarrow \mathbf{x}_{\text{BP}}$ when $z \rightarrow +\infty$.

B. Statement of Main Result

We consider the channel entropy $\mathbf{h} \in [\mathbf{h}_{\text{BP}}, \mathbf{h}_{\text{MAP}}]$. The phenomenology tells us that: (i) after a transient phase, the profile develops a fixed shape $\mathbf{X}(\cdot)$; (ii) the shape is independent of the initial condition; (iii) the shape travels at constant speed v ; (iv) the shape satisfies the boundary conditions $\mathbf{X}(z) \rightarrow \Delta_\infty$ for $z \rightarrow -\infty$ and $\mathbf{X}(z) \rightarrow \mathbf{x}_{\text{BP}}$ for $z \rightarrow +\infty$. We thus formalize these observations and make an ansatz:

Ansatz. For each $\mathbf{h} \in [\mathbf{h}_{\text{BP}}, \mathbf{h}_{\text{MAP}}]$ there exist a constant $v \geq 0$ and a family of probability measures $\mathbf{X}(z)$ (indexed by $z \in \mathbb{R}$) satisfying the boundary conditions $\mathbf{X}(z) \rightarrow \Delta_\infty$ for $z \rightarrow -\infty$ and $\mathbf{X}(z) \rightarrow \mathbf{x}_{\text{BP}}$ for $z \rightarrow +\infty$, such that, for $t \rightarrow +\infty$ and $|z - vt| = O(1)$, the solution of DE (14) is independent of the initial condition and satisfies $\mathbf{x}(z, t) \rightarrow \mathbf{X}(z - vt)$.

Implicit in this ansatz is that we restrict ourselves here to underlying code ensembles that have only one non-trivial (stable) BP fixed point. This is true for regular codes for example (but is not limited to this case). Ensembles with many non-trivial fixed points could lead to more complicated phenomenologies as emphasized in [17] and would require to modify the ansatz.

Velocity of the soliton for general BMS channels. Under the assumption above the velocity of the soliton is given by

$$v = \frac{\Delta E}{\int_{\mathbb{R}} dz \, H\left(\rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)^{\boxtimes 2}\right)}, \quad (15)$$

where ΔE is the *energy gap* defined as

$$\Delta E = W_s(\mathbf{x}_{\text{BP}}) - W_s(\Delta_\infty), \quad (16)$$

and we recall that W_s is the potential (6) of the *uncoupled system*, \mathbf{x}_{BP} is the non-trivial BP fixed point to which the uncoupled system converges, and Δ_∞ is the trivial fixed point (Dirac mass at infinity).

Let us make a few remarks. In this formula, the prime denotes the derivative $\mathbf{X}'(z) = \lim_{\delta \rightarrow 0} \delta^{-1}(\mathbf{X}(z+\delta) - \mathbf{X}(z))$ which is to be interpreted as a difference between two measures. The energy gap is only defined for $\mathbf{h}_{\text{BP}} \leq \mathbf{h} \leq \mathbf{h}_{\text{MAP}}$, that is, when the single potential W_s has a non-trivial non-negative local minimum (see e.g. Figure 1). It is exactly equal to zero when $\mathbf{h} = \mathbf{h}_{\text{MAP}}$, which confirms the fact that the velocity of decoding is zero (no decoding occurs) in this case. Note also that, with our normalizations, $W_s(\Delta_\infty) = 0$.

Formula (15) involves the shape $\mathbf{X}(\cdot)$. Using the DE equation, the ansatz $\mathbf{x}(z, t) \rightarrow \mathbf{X}(z - vt)$, and the approximation $\mathbf{x}(z, t + 1) - \mathbf{x}(z, t) \approx -v\mathbf{X}'(z - vt)$, valid for small v , we find after a change of variables that $\mathbf{X}(z)$ is the solution of

$$\mathbf{X}(z) - v\mathbf{X}'(z) = \int_0^1 du \, c(z - u) \otimes \lambda^{\otimes} \left(\int_0^1 ds \, \rho^{\boxtimes}(\mathbf{X}(z - u + s)) \right). \quad (17)$$

To obtain the shape $\mathbf{X}(z)$ and the velocity v , one must iteratively solve the closed system of equations formed by (15) and (17). Note that the assumption of small v used above is strictly valid for h close to h_{MAP} . However, numerical simulations confirm that in practice the resulting velocity formula is precise over the whole range $[h_{\text{BP}}, h_{\text{MAP}}]$.

IV. DERIVATION OF VELOCITY FORMULA FOR BMS CHANNELS

Let us briefly outline of the main steps of derivation. We first write down a potential functional which gives, in the continuous setting, the DE fixed point equation corresponding to (14). This enables us to formulate the DE iterations as a sort of gradient descent equation (Section IV-A). From there on, we use the ansatz in Section III-B to derive the velocity formula (15).

A. Density evolution as gradient descent

We call $\Delta\mathcal{W}(\mathbf{x})$ the potential functional of the coupled system in the continuum limit obtained from (10). This limit involves an integral over the spatial direction $z \in \mathbb{R}$ and, in order to get a convergent result, we must subtract a “reference energy”. Essentially, any static reference profile, here called $\mathbf{x}_0(z)$, that satisfies the boundary conditions $\mathbf{x}_0(z) \rightarrow \Delta_\infty, z \rightarrow -\infty$ and $\mathbf{x}_0(z) \rightarrow x_{\text{BP}}, z \rightarrow +\infty$, will do the job. For concreteness, one can take a Heaviside-like profile $\mathbf{x}_0(z) = \Delta_\infty, z < 0, \mathbf{x}_0(z) = x_{\text{BP}}, z \geq 0$. The potential functional is thus defined as follows,

$$\Delta\mathcal{W}(\mathbf{x}) = \int_{\mathbb{R}} dz \left(P(z, \mathbf{x}) - P(z, \mathbf{x}_0) \right), \quad (18)$$

where $P(z, \mathbf{x})$ is a z -dependent functional of \mathbf{x} equal to

$$\begin{aligned} P(z, \mathbf{x}) = & \frac{1}{R'(1)} H(R^{\boxtimes}(\mathbf{x}(z, t))) + H(\rho^{\boxtimes}(\mathbf{x}(z, t))) - H(\mathbf{x}(z, t) \boxtimes \rho^{\boxtimes}(\mathbf{x}(z, t))) \\ & - \frac{1}{L'(1)} H\left(c(z) \otimes L^{\otimes} \left(\int_0^1 ds \, \rho^{\boxtimes}(\mathbf{x}(z + s, t)) \right)\right). \end{aligned} \quad (19)$$

In Appendix A, we calculate the functional derivative of $\Delta\mathcal{W}(\mathbf{x})$ in a direction $\eta(z, t)$ defined as

$$\frac{\delta\Delta\mathcal{W}}{\delta\mathbf{x}}[\eta(z, t)] = \frac{d}{d\gamma} \Delta\mathcal{W}(\mathbf{x} + \gamma\eta) \Big|_{\gamma=0}, \quad (20)$$

and find

$$\frac{\delta\Delta\mathcal{W}}{\delta\mathbf{x}}[\eta(z, t)] = \int_{\mathbb{R}} dz \, H \left(\left(\int_0^1 du \, c(z - u) \otimes \lambda^{\otimes} \left(\int_0^1 ds \, \rho^{\boxtimes}(\mathbf{x}(z - u + s, t)) \right) - \mathbf{x}(z, t) \right) \boxtimes \rho^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t) \right) \quad (21)$$

From (14) and (21) we deduce that

$$\int_{\mathbb{R}} dz \, H \left((\mathbf{x}(z, t + 1) - \mathbf{x}(z, t)) \boxtimes \rho^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t) \right) = \frac{\delta\Delta\mathcal{W}}{\delta\mathbf{x}}[\eta(z, t)] \quad (22)$$

We note that, using the duality rule (4) for $\mathbf{a} = \mathbf{x}(z, t+1) - \mathbf{x}(z, t)$ and $b = \rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)$ (recall that η must be a difference of two measures so that b also is such a difference) and the associativity of \boxtimes , this equation can also be formulated as

$$\int_{\mathbb{R}} dz H\left((\mathbf{x}(z, t+1) - \mathbf{x}(z, t)) \otimes (\rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t))\right) = -\frac{\delta \Delta \mathcal{W}}{\delta \mathbf{x}}[\eta(z, t)] \quad (23)$$

In this form, we recognize a sort of infinite-dimensional gradient descent equation in a space of measures. This reformulation of DE forms the basis of the derivation of the velocity formula.

B. Final steps of the derivation

The potential functional can be decomposed in a “single system” part and a contribution that contains the “interaction” due to coupling. We have

$$\Delta \mathcal{W}(\mathbf{x}) = \mathcal{W}_s(\mathbf{x}) + \mathcal{W}_i(\mathbf{x}), \quad (24)$$

with

$$\mathcal{W}_s(\mathbf{x}) = \int_{\mathbb{R}} dz \{P_s(z, \mathbf{x}) - P_s(z, \mathbf{x}_0)\}, \quad (25)$$

$$\mathcal{W}_i(\mathbf{x}) = \int_{\mathbb{R}} dz \{P_i(z, \mathbf{x}) - P_i(z, \mathbf{x}_0)\}, \quad (26)$$

where

$$\begin{aligned} P_s(z, \mathbf{x}) &= \frac{1}{R'(1)} H(R^{\boxtimes}(\mathbf{x}(z, t))) + H(\rho^{\boxtimes}(\mathbf{x}(z, t))) - H(\mathbf{x}(z, t) \boxtimes \rho^{\boxtimes}(\mathbf{x}(z, t))) \\ &\quad - \frac{1}{L'(1)} H(\mathbf{c}(z) \otimes L^{\otimes}(\rho^{\boxtimes}(\mathbf{x}(z, t)))), \end{aligned} \quad (27)$$

and

$$P_i(z, \mathbf{x}) = H\left(\mathbf{c}(z) \otimes L^{\otimes}(\rho^{\boxtimes}(\mathbf{x}(z, t)))\right) - H\left(\mathbf{c}(z) \otimes L^{\otimes}\left(\int_0^1 du \rho^{\boxtimes}(\mathbf{x}(z+u, t))\right)\right). \quad (28)$$

Note, for future use, that in fact $P_s(z, \mathbf{x}) = W_s(\mathbf{x}(z, t))$ is the single system potential (6) “at position z ”. With these definitions, the gradient descent equation (23) can be written as

$$\int_{\mathbb{R}} dz H\left((\mathbf{x}(z, t+1) - \mathbf{x}(z, t)) \boxtimes \rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)\right) = \frac{\delta \mathcal{W}_s}{\delta \mathbf{x}}[\eta(z, t)] + \frac{\delta \mathcal{W}_i}{\delta \mathbf{x}}[\eta(z, t)]. \quad (29)$$

Now, we use the ansatz to compute the three terms in this equation in the regime $t \rightarrow +\infty$, $z \rightarrow +\infty$ such that $|z - vt| = O(1)$. We will choose the direction $\eta(z, t) = \mathbf{X}'(z - vt)$.

We start with the left-hand side of (29). From $\mathbf{x}(z, t) \rightarrow \mathbf{X}(z - vt)$ and the approximation $\mathbf{x}(z, t+1) - \mathbf{x}(z, t) \approx -v\mathbf{X}'(z - vt)$, together with the special choice of $\eta(z, t)$, we can rewrite the left hand side of (29) as

$$v \int_{\mathbb{R}} dz H(\mathbf{X}'(z - vt) \boxtimes \rho'^{\boxtimes}(\mathbf{X}(z - vt)) \boxtimes \mathbf{X}'(z, t)). \quad (30)$$

Using the commutativity of the operator \boxtimes , this is equal to

$$v \int_{\mathbb{R}} dz H(\rho'^{\boxtimes}(\mathbf{X}(z - vt)) \boxtimes \mathbf{X}'(z - vt)^{\boxtimes 2}). \quad (31)$$

Note that we can shift the argument in the integrals $z - vt \rightarrow z$, and this term becomes independent of time.

Now, we consider the first functional derivative on the right hand side of (29), when $\eta(z, t) = \mathbf{X}'(z - vt)$. It should be clear that we can immediately make the change of variables in the integrals $z - vt \rightarrow z$ which simplifies the formulas. By the calculations in Appendix A, we find

$$\frac{\delta \mathcal{W}_s}{\delta \mathbf{X}}[\mathbf{X}'(z)] = \int_{\mathbb{R}} dz \left\{ H(\mathbf{X}(z) \otimes [\rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)]) - H(\mathbf{c} \otimes \lambda^{\otimes}(\rho^{\boxtimes}(\mathbf{X}(z))) \otimes [\rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)]) \right\}. \quad (32)$$

In order to simplify the above, we remark the following

$$\begin{aligned} & \frac{d}{dz} \left\{ \frac{1}{R'(1)} H(R^{\boxtimes}(\mathbf{X}(z))) - H(\mathbf{X}(z) \boxtimes \rho^{\boxtimes}(\mathbf{X}(z))) + H(\rho^{\boxtimes}(\mathbf{X}(z))) - \frac{1}{L'(1)} H(\mathbf{c}(z) \otimes L^{\otimes}(\rho^{\boxtimes}(\mathbf{X}(z)))) \right\} \\ &= H(\rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)) - H(\mathbf{X}'(z) \boxtimes \rho^{\boxtimes}(\mathbf{X}(z))) - H(\mathbf{X}(z) \boxtimes \rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)) + H(\rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)) \\ & \quad - H(\mathbf{c}(z) \otimes \lambda^{\otimes}(\rho^{\boxtimes}(\mathbf{X}(z))) \otimes [\rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)]). \end{aligned}$$

Noticing that the first two terms on the right-hand side cancel out, and using the duality rule (4) for the third term, we get the integrand in (32). In other words, the integrand in (32) equals $\frac{d}{dz} P_s(z, \mathbf{X}) = \frac{d}{dz} W_s(z, \mathbf{X}(z))$ and

$$\begin{aligned} \frac{\delta \mathcal{W}_s}{\delta \mathbf{X}}[\mathbf{X}'(z)] &= \int_{\mathbb{R}} dz \frac{d}{dz} W_s(\mathbf{X}(z)) \\ &= W_s(\mathbf{x}_{\text{BP}}) - W_s(\Delta_{\infty}) \\ &= \Delta E. \end{aligned} \quad (33)$$

We now show that the functional derivative of the interaction part in (29) does not contribute when $\eta(z, t)$ is replaced by $\mathbf{X}'(z)$. By directly applying the definition of the functional derivative, we find

$$\begin{aligned} \frac{\delta \mathcal{W}_i}{\delta \mathbf{X}}[\mathbf{X}'(z)] &= \int_{\mathbb{R}} dz \left\{ H(\mathbf{c} \otimes [\lambda^{\otimes}(\rho^{\boxtimes}(\mathbf{X}(z))) \otimes (\rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z))]) \right. \\ & \quad \left. - H(\mathbf{c} \otimes [\lambda^{\otimes}(\int_0^1 du \rho^{\boxtimes}(\mathbf{X}(z+u))) \otimes (\int_0^1 ds \rho^{\boxtimes}(\mathbf{X}(z+s)) \boxtimes \mathbf{X}'(z+s))]) \right\}. \end{aligned} \quad (34)$$

We notice that the integrand is a total derivative; namely, it is equal to

$$\frac{1}{L'(1)} \frac{d}{dz} \left\{ H(\mathbf{c} \otimes L^{\otimes}(\rho^{\boxtimes}(\mathbf{X}(z)))) - H(\mathbf{c} \otimes L^{\otimes}(\int_0^1 du \rho^{\boxtimes}(\mathbf{X}(z+u)))) \right\}$$

Due to the boundary conditions, we have $\lim_{z \rightarrow -\infty} \mathbf{X}(z) = \lim_{z \rightarrow -\infty} \mathbf{X}(z+u) = \Delta_{\infty}$ and $\lim_{z \rightarrow +\infty} \mathbf{X}(z) = \lim_{z \rightarrow +\infty} \mathbf{X}(z+u) = \mathbf{x}_{\text{BP}}$, and we can conclude that the total derivative integrates to zero, thus

$$\frac{\delta \mathcal{W}_i}{\delta \mathbf{X}}[\mathbf{X}'(z)] = 0. \quad (35)$$

Finally, replacing (31), (33), and (35) in (29) we get the simple relationship

$$v \int_{\mathbb{R}} dz H(\rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)^{\boxtimes 2}) = \Delta E \quad (36)$$

which yields the velocity formula (15).

V. APPLICATIONS TO SPECIFIC CHANNELS AND COMPARISONS WITH NUMERICAL EXPERIMENTS

A. Binary Erasure Channel (BEC)

The formula for the velocity, when transmission takes place over the BEC, can be obtained by directly simplifying the general formula in (15). We note that, since the BEC yields a scalar system, one can also use the formula for

general scalar systems in Section VII (that covers cases beyond coding theory also). We will suppose that the underlying code LDPC(λ, ρ) is such that the DE equation has a single non-trivial fixed point $x_{\text{BP}} \neq 0$. Furthermore, we fix $\epsilon_{\text{BP}} \leq \epsilon \leq \epsilon_{\text{MAP}}$ (recall the channel entropy reduces to $H(c) = h = \epsilon$ here).

The channel distribution can be written as $c = \epsilon\Delta_0 + (1 - \epsilon)\Delta_\infty$, and the profile is of the form $\mathbf{x}(z, t) = x(z, t)\Delta_0 + (1 - x(z, t))\Delta_\infty$ where $0 \leq x(z, t) \leq 1$ is the scalar erasure probability at position z and time t . This tends to a fixed shape

$$\mathbf{X}(z) = X(z)\Delta_0 + (1 - X(z))\Delta_\infty. \quad (37)$$

where $0 \leq X(z) \leq 1$ satisfies $\lim_{z \rightarrow -\infty} X(z) = 0$, $\lim_{z \rightarrow +\infty} X(z) = x_{\text{BP}}$. We have also

$$\mathbf{X}'(z) = X'(z)\Delta_0 - X'(z)\Delta_\infty. \quad (38)$$

We also note the following identities valid for scalar maps $f, g : \mathbb{R} \rightarrow [0, 1]$ (such as λ, ρ, L, R and their derivatives)

$$\begin{cases} f^\oplus(\mathbf{X}(z)) = f(X(z))\Delta_0 + (1 - f(X(z)))\Delta_\infty, \\ g^\boxplus(\mathbf{X}(z)) = (1 - g(1 - X(z)))\Delta_0 + g(1 - X(z))\Delta_\infty. \end{cases} \quad (39)$$

Let us compute the denominator of (15). Using (3), (38), and (39) we have

$$\begin{aligned} \rho'^{\boxplus}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)^{\boxplus 2} &= \{(1 - \rho'(1 - X(z)))\Delta_0 + \rho'(1 - X(z))\Delta_\infty\} \boxtimes \{X'(z)\Delta_0 - X'(z)\Delta_\infty\}^{\boxplus 2} \\ &= \{(1 - \rho'(1 - X(z)))\Delta_0 + \rho'(1 - X(z))\Delta_\infty\} \boxtimes \{X'(z)^2\Delta_\infty - X'(z)^2\Delta_0\} \\ &= (1 - \rho'(1 - X(z)))X'(z)^2\Delta_0 - (1 - \rho'(1 - X(z)))X'(z)^2\Delta_0 \\ &\quad + \rho'(1 - X(z))X'(z)^2\Delta_\infty - \rho'(1 - X(z))X'(z)^2\Delta_0 \\ &= \rho'(1 - X(z))X'(z)^2\Delta_\infty - \rho'(1 - X(z))X'(z)^2\Delta_0, \end{aligned}$$

and since $H(\Delta_0) = 1$, $H(\Delta_\infty) = 0$, and the entropy functional is linear, we obtain the denominator of (15) as

$$H(\rho'^{\boxplus}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)^{\boxplus 2}) = -\rho'(1 - X(z))X'(z)^2.$$

For the numerator of (15), we have $\Delta E = W_{\text{BEC}}(x_{\text{BP}}) - W_{\text{BEC}}(0)$, where the single system potential on the BEC is obtained from (6) using again (3) and (39). The exercise yields

$$W_{\text{BEC}}(x) = \frac{1}{R'(1)}(1 - R(1 - x)) - x\rho(1 - x) - \frac{\epsilon}{L'(1)}L(1 - \rho(1 - x)).$$

Putting together these results, the velocity (15) becomes

$$v_{\text{BEC}} = -\frac{W_{\text{BEC}}(x_{\text{BP}}) - W_{\text{BEC}}(0)}{\int_{\mathbb{R}} dz \rho'(1 - X(z))X'(z)^2}. \quad (40)$$

(Note that, with our normalizations, $W_{\text{BEC}}(0) = 0$ for all ϵ .) The erasure profile $X(z)$ has to be computed from the one-dimensional integral equation

$$X(z) - v_{\text{BEC}}X'(z) = \epsilon \int_0^1 du \lambda\left(1 - \int_0^1 ds \rho(1 - X(z - u + s))\right). \quad (41)$$

Obviously, the velocity vanishes when $\epsilon \rightarrow \epsilon_{\text{MAP}}$ since then $W_{\text{BEC}}(x_{\text{BP}}) \rightarrow W_{\text{BEC}}(0) = 0$. An important quantity is the slope of the velocity at ϵ_{MAP} . To compute it, we remark that W_{BEC} has an explicit dependence on ϵ , as well as an implicit one through $x_{\text{BP}}(\epsilon)$. Thus,

$$\begin{aligned} \frac{dW_{\text{BEC}}}{d\epsilon} &= \frac{\partial W_{\text{BEC}}}{\partial \epsilon} + \frac{\partial W_{\text{BEC}}}{\partial x_{\text{BP}}} \frac{dx_{\text{BP}}}{d\epsilon} \\ &= \frac{\partial W_{\text{BEC}}}{\partial \epsilon} \\ &= -\frac{1}{L'(1)} L(1 - \rho(1 - x_{\text{BP}})), \end{aligned}$$

so that, for $\epsilon \rightarrow \epsilon_{\text{MAP}}$, the Taylor expansion to first order yields

$$W_{\text{BEC}}(x_{\text{BP}}) \approx -(\epsilon - \epsilon_{\text{MAP}}) \frac{1}{L'(1)} L(1 - \rho(1 - x_{\text{MAP}})).$$

Note that we used $x_{\text{BP}}(\epsilon) \rightarrow x_{\text{BP}}(\epsilon_{\text{MAP}}) = x_{\text{MAP}}$ where x_{MAP} is defined as the point $x \neq 0$ where the potential is stationary and vanishes. This yields the linear approximation for the velocity

$$v_l = \frac{(\epsilon - \epsilon_{\text{MAP}})}{L'(1)} \frac{L(1 - \rho(1 - x_{\text{MAP}}))}{\int_{\mathbb{R}} dz \rho'(1 - X_{\text{MAP}}(z))(X'_{\text{MAP}}(z))^2}, \quad (42)$$

where $X_{\text{MAP}}(\cdot)$ is the erasure probability profile obtained when $\epsilon = \epsilon_{\text{MAP}}$.

It is interesting to compare (40) with the upper bound of Theorem 1 in [17] for a *discrete* system

$$v_B = \alpha \frac{W_{\text{BEC}}(x_{\text{BP}}) - W_{\text{BEC}}(0)}{\sum_{z \in \mathbb{Z}} \rho'(1 - x_z)(x_z - x_{z-1})^2}, \quad \alpha \leq 2. \quad (43)$$

In [17] the derivation of the bound yields $\alpha \leq 2$ (for L_c and w large enough) but it is conjectured based on numerical simulations that $\alpha = 1$ would be a tight bound. Obviously, (40) and (43) are consistent. We note for reference that another upper bound is also derived in [17], namely

$$v_{B_2} = \frac{\alpha(W_{\text{BEC}}(x_{\text{BP}}; \epsilon) - W_{\text{BEC}}(0; \epsilon))}{2W_{\text{BP}}(x_u; \epsilon) - W_{\text{BP}}(x_{\text{BP}}; \epsilon)},$$

where x_u and x_{BP} are respectively the non-trivial unstable and stable fixed points of the potential of the uncoupled system $W_s(\cdot; \cdot)$. We do not discuss this further because in practice this turns out to be a very loose bound.

We now compare the *analytical* velocity formula (40) with the *empirical* velocity (called v_e below) obtained by simulating the discrete DE equation; we show that it provides a very good approximation for the (real) value of the velocity even for relatively small values of w . For the simulations, we consider the spatially coupled (3, 6) and (4, 6)-regular code ensembles, as well as two irregular LDPC codes (described later). We run the simulations for several values of the chain length $L_c = 256, 1024$ and the window size $w = 3, 5, 8, 16$. The empirical velocity is the velocity calculated from erasure probability profiles of the discrete DE equation. Consider two (discrete) profiles $\underline{x}^{(t_1)}$ and $\underline{x}^{(t_2)}$ at any two iterations t_1 and t_2 , respectively, with $t_1 < t_2$. After the transient phase is over, the profiles are identical up to translation. We call a “kink” the part of the profile where there is a fast increase from 0 to x_{BP} in the erasure probability. The kink “position” is the coordinate such that the height is equal to $x_{\text{BP}}/2$, and Δz is the difference of two such positions (on two different profiles). Then, the empirical velocity v_e is

$$v_e = \frac{\Delta z}{w(t_2 - t_1)} \quad (44)$$

In practice, we get reliable results by taking pairs of profiles separated by 20 iterations and averaging this ratio over every consecutive pair of profiles. Note that we normalize the velocity by w to be able to compare systems with different window widths.

In Table I, we give empirical values v_e of the normalized velocities for the spatially coupled $(4, 6)$ -regular code ensemble, with transmission over the BEC(0.6), when the spatial length is 1024 positions and the channel parameter is fixed to $\epsilon = 0.6$ (between the BP and MAP thresholds), for different values of the window size w . We observe that the result of our formula v_{BEC} gives a good estimate of the empirical velocity v_e for all the demonstrated values of the window size. We also observe that the linear approximation gives a good estimate when the channel parameter is not too far from the MAP threshold ϵ_{MAP} . The upper bound v_B [17] gives a better estimate as the window size grows larger.

TABLE I

NORMALIZED VELOCITIES FOR LDPC(x^4, x^6) ON THE BEC WITH A SPATIAL LENGTH 1024, FOR SEVERAL w SIZES, AND $\epsilon = 0.6$. THE VALUES IN THE TABLE CAN BE COMPARED TO $v_{\text{BEC}} = 0.0333$ AND $v_l = 0.0293$.

	$w = 3$	$w = 5$	$w = 8$	$w = 16$
v_e	0.0325	0.0335	0.0337	0.0339
v_B/α	0.0473	0.0410	0.0380	0.0356

In Table II, we give empirical values of the normalized velocities for the spatially coupled $(3, 6)$ -regular code ensemble, with transmission over the BEC(ϵ), when the spatial length equals 1024 and the window size $w = 8$, for different values of the channel parameter ϵ . One can compare these values with those in [17] (up to a factor equal to w due to the normalization). The result of the formula v_{BEC} gives the closest estimate to the empirical velocity v_e for all values of ϵ .

TABLE II

NORMALIZED VELOCITIES FOR LDPC(x^3, x^6) ON THE BEC FOR SPATIAL LENGTH 1024, $w = 8$, AND SEVERAL ϵ VALUES.

	$\epsilon = 0.45$	$\epsilon = 0.46$	$\epsilon = 0.47$	$\epsilon = 0.48$
v_e	0.0667	0.0458	0.0267	0.0117
v_{BEC}	0.0660	0.0449	0.0272	0.0115
v_l	0.0506	0.0373	0.0240	0.0108
v_B/α	0.0781	0.0541	0.0332	0.0142
v_{B_2}/α	0.6970	0.5008	0.3068	0.1291

Figures 4 and 5 show the empirical velocity v_e , the analytical velocity v_{BEC} , and the upper bound v_B for the spatially coupled $(3, 6)$ -regular code ensemble, with spatial length 256 and window size $w = 3$. We remark that our formula fits very well, for the $(3, 6)$ -regular code, with the empirical velocity for all values of the channel

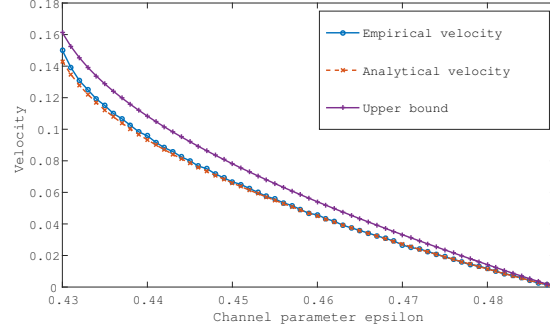


Fig. 4. Normalized velocities v_e , v_{BEC} , and v_B/α (in the order of the legend) for the $(3,6)$ -regular ensemble with spatial length 256, window size $w = 3$, and $\epsilon_{\text{BP}} = 0.43 < \epsilon < \epsilon_{\text{MAP}} = 0.4881$.

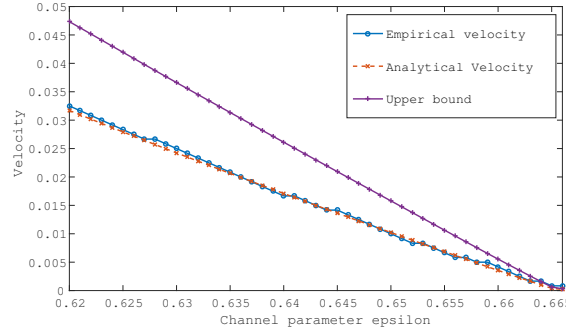


Fig. 5. Normalized velocities v_e , v_{BEC} , and v_B/α (in the order of the legend) of the decoding profile for the $(4,6)$ -regular ensemble of spatial length 256, window size $w = 3$, and $\epsilon_{\text{BP}} = 0.515 < \epsilon < \epsilon_{\text{MAP}} = 0.666$.

parameter $\epsilon \in [\epsilon_{\text{BP}}, \epsilon_{\text{MAP}}] = [0.43, 0.488]$. The agreement is quite good also for the $(4,6)$ -regular code and very good for more than half of the interval $[\epsilon_{\text{BP}}, \epsilon_{\text{MAP}}] \approx [0.515, 0.719]$.

We also illustrate the results for two irregular code ensembles in Figures 6 and 7. The first one has node degree distributions $L(x) = 0.3x^2 + 0.6x^3 + 0.1x^5$ and $R(x) = x^4$, spatial length 1024, and window size $w = 4$. The agreement between v_{BEC} and v_e is excellent for the whole range $\epsilon \in [\epsilon_{\text{BP}}, \epsilon_{\text{MAP}}] = [0.657, 0.719]$. The second one has $L(x) = 0.4x^3 + 0.3x^4 + 0.3x^5$ and $R(x) = 0.5x^8 + 0.5x^{12}$, spatial length 256, and window size $w = 3$. The agreement between the velocities is also very good for most of the range $\epsilon \in [\epsilon_{\text{BP}}, \epsilon_{\text{MAP}}] = [0.311, 0.385]$.

B. Gaussian Approximation (GA)

DE equations relate probability densities and as such we may need to track an infinite set of parameters (except for the BEC where the space of densities can be parametrized by a single real number). In many situations, such as the case when we have large degrees, for example, the densities are well approximated by Gaussians, which enables us to project the DE equations down to a low dimensional space. There are several variants of the Gaussian approximation (see for example [31], [18], [19]), and here we use it in a form called the “reciprocal channel approximation” proposed in [18], [19].

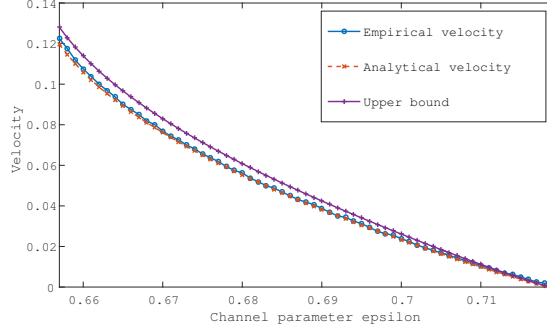


Fig. 6. Normalized velocities v_{BEC} , v_e , and v_B/α for an ensemble with $L(x) = 0.3x^2 + 0.6x^3 + 0.1x^5$, $R(x) = x^4$, spatial length 1024, window size $w = 4$, and $\epsilon_{\text{BP}} = 0.657 < \epsilon < \epsilon_{\text{MAP}} = 0.719$.

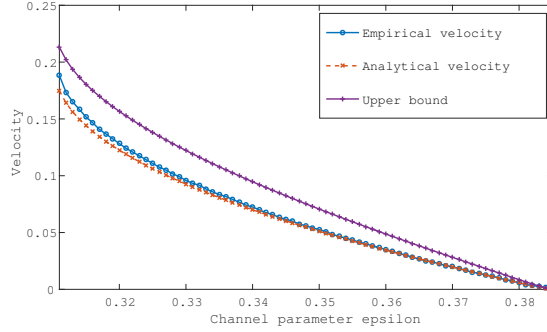


Fig. 7. Normalized velocities v_{BEC} , v_e , v_B/α for an ensemble with $L(x) = 0.4x^3 + 0.3x^4 + 0.3x^5$, $R(x) = 0.5x^8 + 0.5x^{12}$, spatial length 256, window size $w = 3$, and $\epsilon_{\text{BP}} = 0.311 < \epsilon < \epsilon_{\text{MAP}} = 0.385$.

The idea is to assume that the densities of the LLR messages appearing in the DE equations are symmetric Gaussian densities. Such densities take the form

$$d\mathbf{x}(\alpha) = \frac{d\alpha}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\alpha - m)^2}{2\sigma^2}\right), \quad (45)$$

with the mean m and variance σ^2 satisfying $\sigma^2 = 2m$. Furthermore, the channel density c is replaced by that corresponding to a BIAWGNC(σ_n^2) with the same entropy $H(c)$. Density evolution can then conveniently be expressed in terms of the entropies $p_z^{(t)} = H(\mathbf{x}_z^{(t)})$. This is done as follows. Let $\psi(m)$ denote the entropy² of a symmetric Gaussian density of mean m given by

$$\psi(m) = \frac{1}{\sqrt{4\pi m}} \int_{\mathbb{R}} dz e^{-\frac{(z-m)^2}{4m}} \log_2(1 + e^{-z}). \quad (46)$$

Thus $\psi^{-1}(p)$ denotes the mean of a symmetric Gaussian density \mathbf{x} of entropy $p = H(\mathbf{x})$. Take two symmetric Gaussian densities \mathbf{x}_1 and \mathbf{x}_2 of means m_1 and m_2 and entropies $p_1 = \psi(m_1)$ and $p_2 = \psi(m_2)$. We have, in

²For indications on the numerical implementation of this function see [27], pp.194 and 237.

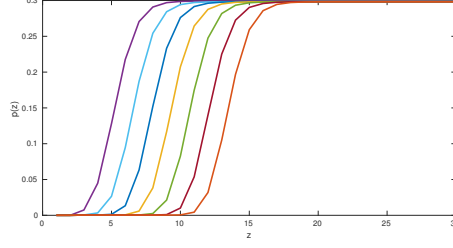


Fig. 8. The profile of entropies $p(z, t)$ plotted every 10 iterations starting from iteration 20. We take the $(3, 6)$ -regular LDPC code with chain length $L_c = 30$, window size $w = 3$, and we consider the BIAWGN channel with mean $\psi^{-1}(H(\mathbf{c})) = 2.4$.

general,

$$H(\mathbf{x}_1 \otimes \mathbf{x}_2) = \psi(\psi^{-1}(p_1) + \psi^{-1}(p_2)), \quad (47)$$

which just expresses the fact that a usual convolution of two Gaussian densities of means m_1 and m_2 is a Gaussian density of mean $m_1 + m_2$. On the other hand $\mathbf{x}_1 \boxtimes \mathbf{x}_2$ is not exactly Gaussian so there is no exact formula but the idea here is to preserve the duality rule $H(\mathbf{x}_1 \boxtimes \mathbf{x}_2) + H(\mathbf{x}_1 \otimes \mathbf{x}_2) = H(\mathbf{x}_1) + H(\mathbf{x}_2)$. Writing this relation as

$$1 - H(\mathbf{x}_1 \boxtimes \mathbf{x}_2) = H((\Delta_0 - \mathbf{x}_1) \otimes (\Delta_0 - \mathbf{x}_2)),$$

and noting that $H(\Delta_0 - \mathbf{x}_1) = 1 - p_1$, $H(\Delta_0 - \mathbf{x}_2) = 1 - p_2$ suggests the approximation

$$H(\mathbf{x}_1 \boxtimes \mathbf{x}_2) = 1 - \psi(\psi^{-1}(1 - p_1) + \psi^{-1}(1 - p_2)). \quad (48)$$

Looking at the entropies of the DE equations (47) and (48) imply (we will limit ourselves to regular codes for simplicity)

$$\begin{cases} H(\mathbf{x}^{(t)}) = H(\mathbf{c} \otimes \mathbf{y}^{(t) \otimes (\ell-1)}), \\ H(\mathbf{y}^{(t+1)}) = H(\mathbf{x}^{(t) \boxtimes (r-1)}). \end{cases} \quad (49)$$

and setting $p^{(t)} = H(\mathbf{x}^{(t)})$, $q^{(t)} = H(\mathbf{y}^{(t)})$ we find

$$\begin{cases} p^{(t)} = \psi(\psi^{-1}(H(\mathbf{c})) + (\ell - 1)\psi^{-1}(q^{(t)})), \\ q^{(t+1)} = 1 - \psi((r - 1)\psi^{-1}(1 - p^{(t)})). \end{cases} \quad (50)$$

These equations can be combined into

$$p^{(t+1)} = \psi\left((\ell - 1)\psi^{-1}(1 - \psi((r - 1)\psi^{-1}(1 - p^{(t)}))) + \psi^{-1}(H(\mathbf{c}))\right). \quad (51)$$

The corresponding potential function is easily found from (6)

$$\begin{aligned} W_{\text{GA}}(p) = & \frac{1}{r} (1 - \psi(r\psi^{-1}(1 - p))) + \psi(r\psi^{-1}(1 - p)) - \psi((r - 1)\psi^{-1}(1 - p)) \\ & - \frac{1}{\ell} \psi(\psi^{-1}(H(\mathbf{c})) + \ell\psi^{-1}(1 - \psi((r - 1)\psi^{-1}(1 - p)))). \end{aligned} \quad (52)$$

For the coupled system, we denote by p_z the *average over positions* $\{z, \dots, z+w\}$ of the entropy of symmetric Gaussian densities emanating from the variable nodes. The coupled DE equations then take the form

$$p_z^{(t+1)} = \frac{1}{w} \sum_{i=0}^{w-1} \psi \left((\ell-1) \psi^{-1} \left(\frac{1}{w} \sum_{j=0}^{w-1} (1 - \psi((r-1) \psi^{-1}(1 - p_{z-i+j}^{(t)}))) \right) + \psi^{-1}(H(c)) \right). \quad (53)$$

This coupled recursion can be solved with appropriate boundary conditions and one observes a scalar wave propagation, as shown in Figure 8.

We are now ready to discuss the application to the velocity formula. The continuum limit is obtained exactly as in Section III-A. The assumption that the density $\mathbf{x}(z, t)$ tends to a fixed shape $\mathbf{X}(z - v_{\text{GA}} t)$ after the transient phase implies that its entropy $p(z, t)$ tends to $P(z - v_{\text{GA}} t) \equiv H(\mathbf{X}(z - v_{\text{GA}} t))$, where $P(z)$ is a scalar function (independent of initial conditions) satisfying the integral equation

$$P(z) - v_{\text{GA}} P'(z) = \int_0^1 du \psi \left(\psi^{-1}(H(c)) + (\ell-1) \psi^{-1} \left(1 - \int_0^1 ds \psi((r-1) \psi^{-1}(1 - P(z - u + s))) \right) \right). \quad (54)$$

and the boundary conditions $\lim_{z \rightarrow -\infty} P(z) = 0$ and $\lim_{z \rightarrow +\infty} P(z) = p_{\text{BP}}$ where p_{BP} is the non-trivial fixed point of (51). We will now show that the velocity formula reduces to

$$v_{\text{GA}} = - \frac{W_{\text{GA}}(p_{\text{BP}}) - W_{\text{GA}}(0)}{(r-1) \int_{\mathbb{R}} dz P'(z)^2 \frac{\psi''((r-2)\psi^{-1}(1-P(z)))}{\psi'(\psi^{-1}(1-P(z)))^2}}. \quad (55)$$

To derive (55), we consider the denominator in (15) and write it as follows

$$\begin{aligned} (r-1) H(\mathbf{x}(z)^{\boxtimes(r-2)} \boxtimes \mathbf{x}'(z)^{\boxtimes 2}) &= (r-1) \lim_{\delta \rightarrow 0} \frac{1}{\delta^2} \left\{ H(\mathbf{x}(z)^{\boxtimes(r-2)} \boxtimes \mathbf{x}(z+\delta)^{\boxtimes 2}) \right. \\ &\quad \left. - 2H(\mathbf{x}(z)^{\boxtimes(r-2)} \boxtimes \mathbf{x}(z+\delta) \boxtimes \mathbf{x}(z)) + H(\mathbf{x}(z)^{\boxtimes(r-2)} \boxtimes \mathbf{x}(z)^{\boxtimes 2}) \right\}. \end{aligned} \quad (56)$$

Computing each entropy in the Gaussian approximation, we find for the bracket on the right-hand side

$$\begin{aligned} \frac{1}{\delta^2} \left\{ [1 - \psi((r-2)\psi^{-1}(1-p(z)) + 2\psi^{-1}(1-p(z+\delta)))] \right. \\ \left. - 2[1 - \psi((r-1)\psi^{-1}(1-p(z)) + \psi^{-1}(1-p(z+\delta)))] + [1 - \psi(r\psi^{-1}(1-p(z)))] \right\}. \end{aligned} \quad (57)$$

In Appendix B, we compute the limit of this term when $\delta \rightarrow 0$ by appropriate Taylor expansions and find

$$- \left(\frac{dP(z)}{dz} \right)^2 \frac{\psi''((r-2)\psi^{-1}(1-P(z)))}{(\psi'(\psi^{-1}(1-P(z))))^2}. \quad (58)$$

This concludes the derivation of (55).

Table III gives a comparison of analytical and empirical velocities, v_{GA} and $v_{e,\text{GA}}$, obtained for the (3, 6) and the (4, 8)-regular ensembles, for a spatial length of 100 and a window size $w = 3$ for different values of $\psi^{-1}(H(c)) = \sigma_n^2/2$ (twice the signal to noise ratio). We also plot both velocities for the (3, 6)-regular ensemble for the same parameters in Figure 9. We conjecture that the errors incurred from these plots are due to numerical errors involved in computing the functions ψ and its inverse.

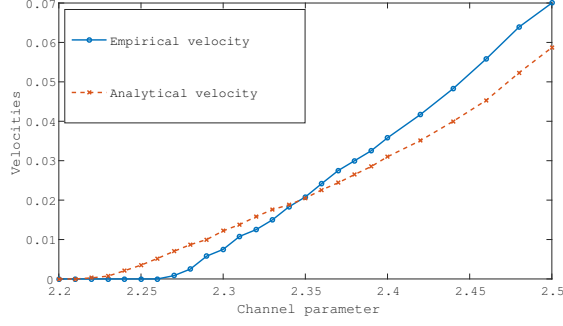


Fig. 9. Normalized velocities v_{GA} and v_e for the spatially coupled $(3, 6)$ -regular ensemble within the Gaussian approximation framework, for a spatial length of 100 and $w = 3$, as a function of $\psi^{-1}(H(c)) = 2/\sigma_n^2$.

TABLE III

NORMALIZED VELOCITIES OF THE WAVES WITHIN THE GAUSSIAN APPROXIMATION ON THE $(3, 6)$ AND $(4, 8)$ -REGULAR CODE ENSEMBLES WITH $L_c = 100$, $w = 3$.

$2/\sigma_n^2$	2.33	2.35	2.38	2.40
$v_{\text{GA}} (3, 6)$	0.0176	0.0205	0.0265	0.0310
$v_e (3, 6)$	0.0150	0.0208	0.0300	0.0358
$v_{\text{GA}} (4, 8)$	0.0237	0.0258	0.0312	0.0381
$v_e (4, 8)$	0.0217	0.0250	0.0308	0.0342

VI. APPLICATION TO SCALING LAWS FOR FINITE-LENGTH COUPLED CODES

The authors in [20] propose a scaling law to predict the error probability of a finite-length spatially coupled (ℓ, r, L_c) code when transmission takes place over the BEC. The derived scaling law depends on *scaling parameters*, one of which we will relate to the velocity of the decoding wave. The (ℓ, r, L_c) ensemble considered in [20] differs slightly from the purely random ensemble we consider in this work. However, as we will see, our formula for the velocity yields results that are reasonably good for this application. We briefly describe this ensemble and the scaling law.

The (ℓ, r, L_c) ensemble combines the benefits of purely random codes (that we consider in this work) and protograph-based codes [32]. The randomness involved in the construction makes the ensemble relatively easy to analyze, and the structure added to the construction due to its similarity to protograph-based codes improves the performance of the code. The ensemble is constructed as follows: Make $L_c + w$ copies of an uncoupled code at positions $z = -w + 1, \dots, L_c$. All edges are erased then reconnected such that a variable node at position z_0 has exactly one edge with each set of check nodes at positions $z_0 + i$, where $i = 0, \dots, \ell - 1$. The check nodes are chosen such that the regularity of their degree is maintained. Therefore, every variable node has ℓ emanating edges and every check node has r such edges.

We consider transmission over the BEC. In this case, the BP decoder can be seen as a peeling decoder [33].

Whenever a variable node is decoded, it is removed from the graph along with its edges. One way to track this peeling process is to analyze the evolution of the degree distribution of the residual graph across iterations, which serves as a sufficient statistic. This statistic can be described by a system of differential equations, whose solution determines the mean and variance of the *fraction of degree-one check nodes* and the variance around this mean at any time during the decoding process. We call \hat{r}_1 the mean.

It has been shown in [20] that there exists a *steady state phase* where the mean and the variance are constant. It is exactly during this phase that one can observe the progression of the soliton. We note that here we consider one-sided termination instead of two-sided termination (as considered in [20]), so the fraction \hat{r}_1 here is equal to half the fraction called $\hat{r}_1(*)$ in [20].

Let $\epsilon_{(\ell,r,L_c)}$ denote the BP threshold of the finite-size (ℓ, r, L_c) ensemble (for large L_c this is close to ϵ_{MAP} due to threshold saturation). We can write the first-order Taylor expansion of $\hat{r}_1|_\epsilon$ around $\epsilon < \epsilon_{(\ell,r,L_c)}$ as

$$\hat{r}_1|_\epsilon \approx \hat{r}_1|_{\epsilon_{(\ell,r,L_c)}} + \gamma \Delta\epsilon + \mathcal{O}(\Delta\epsilon^2).$$

where $\Delta\epsilon = \epsilon_{(\ell,r,L_c)} - \epsilon$. Thus, for a given $\epsilon < \epsilon_{(\ell,r,L_c)}$ and since $\hat{r}_1|_{\epsilon_{(\ell,r,L_c)}} = 0$ (by definition), we obtain

$$\gamma \approx \frac{\hat{r}_1|_\epsilon}{\Delta\epsilon}.$$

This parameter γ enters in the scaling law and is therefore quite important. So far, it has been determined only experimentally. It would clearly be desirable to have a theoretical handle on γ . It is argued in [20] that $\gamma \approx \bar{\gamma}$ where $\bar{\gamma} = x_{\text{BP}}/c$ and c is a real positive constant that behaves like $\Delta\epsilon/v_{\text{BEC}}$, i.e.,

$$\bar{\gamma} \approx \frac{x_{\text{BP}} v_{\text{BEC}}}{\Delta\epsilon} \quad (59)$$

It is expected that this formula becomes exact in an asymptotic limit where threshold saturation takes place $\epsilon_{(\ell,r,L_c)} \rightarrow \epsilon_{\text{MAP}}$. Using the linearization (42), we obtain

$$\bar{\gamma} \rightarrow \frac{x_{\text{MAP}}}{L'(1)} \frac{L(1 - \rho(1 - x_{\text{MAP}}))}{\int_{\mathbb{R}} dz \rho'(1 - X_{\text{MAP}}(z))(X'_{\text{MAP}}(z))^2}. \quad (60)$$

The parameter $\bar{\gamma}$ is simply equal to the erasure probability times the slope of the velocity at ϵ_{MAP} .

We compare the values of γ and $\bar{\gamma}$ for different values of ℓ and r , at a channel parameter $\epsilon = \epsilon_{(\ell,r,L_c)} - 0.04$, in Table IV. The experimental values of γ are taken from [20] and, for those of $\bar{\gamma}$, we use the analytical velocity (40). We observe that the numbers roughly agree. There are two reasons that can explain the discrepancies. Firstly, we derive the velocity for the purely random spatially coupled graph ensemble whereas the ensemble considered in [20] is more structured. Note also that as ℓ, r increase, the window size of the structured ensemble increases, so the finite size effects at fixed $L_c = 100$ may be more marked. Secondly, expression (59) is valid when $\epsilon \rightarrow \epsilon_{(\ell,r,L_c)}$, whereas in IV $\Delta\epsilon = 0.04$, which is relatively large (this choice in [20] is due to stability issues in numerical integration techniques when $\epsilon \rightarrow \epsilon_{(\ell,r,L_c)}$). We conjecture that the second issue is the dominant reason for the difference between the values of γ and $\bar{\gamma}$ and that, in fact, the velocity for the structured ensemble is not very different from the one predicted by our formula (40).

TABLE IV
VALUES OF γ AND $\bar{\gamma}$ FOR $\epsilon_{(\ell,r,L)} - \epsilon = 0.04$, $L_c = 100$, AND SEVERAL VALUES OF ℓ AND r

l	r	ϵ_{MAP}	γ	$\bar{\gamma}$
3	6	0.4881	2.155	1.960
4	8	0.4977	2.120	1.779
5	10	0.4994	2.095	1.733
6	12	0.4999	2.075	1.722
4	12	0.3302	2.140	1.778
5	15	0.3325	2.115	1.746
4	6	0.6656	2.100	1.735

VII. VELOCITY FOR GENERAL SCALAR COUPLED SYSTEMS

In this section, we consider general scalar spatially coupled systems. That is, we do not restrict ourselves to coding problems; however, we only consider systems in which the “density evolution type” analysis of message passing algorithms involves *scalar* values. Our main result is again a general formula for the velocity of the soliton in the framework of general scalar coupled systems. There are numerous systems that fall in this class - coding with transmission on the BEC being one of the simplest - and in Sections VII-D, VII-E we will illustrate our results with two examples, namely generalized LDPC codes on the BEC and BSC, as well as compressive sensing.

A. General scalar systems

We adopt the framework and notation in [21]. We denote by $\epsilon \in [0, \epsilon_{\max}]$ where $\epsilon_{\max} \in (0, \infty)$ the interval of values for the control parameter ϵ . Consider bounded, smooth functions that are increasing in both their arguments $g : [0, x_{\max}(\epsilon)] \times [0, \epsilon_{\max}] \rightarrow [0, y_{\max}(\epsilon)]$ and $f : [0, y_{\max}(\epsilon)] \times [0, \epsilon_{\max}] \rightarrow [0, x_{\max}(\epsilon)]$ where $x_{\max}(\epsilon), y_{\max}(\epsilon) \in (0, \infty)$ and $y_{\max}(\epsilon) = g(x_{\max}(\epsilon); \epsilon)$. The scalar recursions that interest us are

$$x^{(t+1)} = f(g(x^{(t)}; \epsilon); \epsilon), \quad (61)$$

where $t \in \mathbb{N}$ is the iteration number. The recursion is initialized with $x^{(0)} = x_{\max}$. Since $f(g([0, x_{\max}(\epsilon)])) \subset [0, x_{\max}(\epsilon)]$, the initialization of (61) implies that $x^{(1)} \leq x^{(0)} = x_{\max}$ and more generally $x^{(t+1)} \leq x^{(t)}$. Thus $x^{(t)}$ will converge to a limiting value $x^{(\infty)}$ and this limit is a fixed point since f and g are continuous. The fixed points of the recursion (61) can be described as stationary points of a *single system* potential function U_s defined as

$$U_s(x) = xg(x; \epsilon) - G(x; \epsilon) - F(g(x; \epsilon); \epsilon), \quad (62)$$

where $F(x; \epsilon) = \int_{g(0; \epsilon)}^x ds f(s; \epsilon)$ and $G(x; \epsilon) = \int_0^x ds g(s; \epsilon)$. Without loss of generality, this function is normalized so that $U_s(x) = 0$.

We define $x_{\text{good}}(\epsilon)$ as the fixed point of the recursion (61) that is reached with an initialization $x^{(0)} = 0$. Furthermore, the *algorithmic threshold*³ is defined as

$$\epsilon_a = \sup\{\epsilon \mid x^{(\infty)} = x_{\text{good}}\}. \quad (63)$$

The monotonicity of f and g implies that, for $\epsilon < \epsilon_a$, the basin of attraction of x_{good} is the whole interval $[0, x_{\text{max}}(\epsilon)]$. Moreover x_{good} is the unique stationary point of the potential function and is a minimum since it is an attractive fixed point. For $\epsilon > \epsilon_a$ we have $x^{(\infty)} \neq x_{\text{good}}$ and we set $x^{(\infty)} = x_{\text{bad}}$ (where x_{bad} depends on ϵ). Note that this is an attractive fixed point and is thus a (local) minimum of $U_s(x)$. The two attractive fixed points are separated by at least one unstable fixed point x_{unst} which is a local maximum of $U_s(x)$. We henceforth assume that there does not appear any other fixed point besides x_{good} , x_{unst} , and x_{bad} . With this assumption in mind, we define the *energy gap* as

$$\Delta E = U_s(x_{\text{bad}}) - U_s(x_{\text{good}}). \quad (64)$$

and the *potential threshold* as the unique value ϵ_{pot} such that $\Delta E = 0$ (it can be shown that ΔE is non-increasing in ϵ).

The corresponding spatially coupled recursions are obtained by placing $L_c + w$ replicas of the single system on the spatial positions $z \in \{-w + 1, \dots, L_c\}$ and coupling them with a uniform coupling window of size w . The coupled recursion takes the form

$$x_z^{(t+1)} = \frac{1}{w} \sum_{j=0}^{w-1} f\left(\frac{1}{w} \sum_{k=0}^{w-1} g(x_{z-j+k}^{(t)}; \epsilon); \epsilon\right). \quad (65)$$

Motivated by the phenomenology observed in many examples (e.g. for the BEC or for compressive sensing), in order to study the stationary phase where a soliton appears, we fix the boundary conditions as $x_z^{(t)} = x_{\text{good}}$, for $z = \{-w + 1, \dots, -1\}$ and all $t \in \mathbb{N}$. The initialization of the recursion is $x_z^{(0)} = x_{\text{bad}}$, for $z = \{0, \dots, L_c\}$. The corresponding *potential functional* is given by

$$U(\mathbf{x}) = \sum_{z=-w+1}^{L_c} (x_z g(x_z; \epsilon) - G(x_z; \epsilon)) - \sum_{z=-w+1}^{L_c} F\left(\frac{1}{w} \sum_{i=0}^{w-1} g(x_{z+i}; \epsilon); \epsilon\right), \quad (66)$$

where $\mathbf{x} = (x_{-w+1}, \dots, x_{L_c})$. The fixed point equation (65) can be obtained by setting the derivative with respect to \mathbf{x} of the potential $U_c(\mathbf{x})$ to zero.

The spatially coupled recursions (65) display the *threshold saturation* property. Namely, for all $\epsilon < \epsilon_{\text{pot}}$ the fixed point $x_z^{(\infty)}$, $z = -w + 1, \dots, L$, of the recursion (65) is equal to a constant profile x_{good} . In the remainder of this section, we consider the range $\epsilon \in [\epsilon_{\text{BP}}, \epsilon_{\text{MAP}}]$. It is for these values of the parameter ϵ that a soliton propagating at finite speed is observed, after a transient phase lasting only for a few iterations. The soliton is a kink with a front at position z_{front} , making a quick transition of width $O(2w)$, between the two values $x_z^{(t)} \approx x_{\text{good}}$ for $z \ll z_{\text{front}}$ and $x_z^{(t)} \approx x_{\text{bad}}$ for $z \gg z_{\text{front}}$.

The simplest example to keep in mind for the setting described above, as well as for the next paragraph, is the case of LDPC(λ, ρ) codes with transmission over the BEC(ϵ) where $f(x; \epsilon) = \epsilon \lambda(x)$ and $g(x; \epsilon) = \rho(x)$ and $U_s(x)$ is equal to (12). Here $\epsilon_a = \epsilon_{\text{BP}}$, $\epsilon_{\text{pot}} = \epsilon_{\text{MAP}}$, $x_{\text{good}} = 0$ and $x_{\text{bad}} = x_{\text{BP}}$ is the non-trivial BP fixed point.

³Here, we mean the algorithmic threshold of the message passing algorithm underlying the recursion.

B. Continuum limit and velocity formula for scalar systems

We consider the system in the limit $L_c \gg w \gg 1$ and formulate a continuum approximation. The coupled recursion (65) becomes

$$x(z, t+1) = \int_0^1 du f\left(\int_0^1 ds g(x(z-u+s, t); \epsilon); \epsilon\right). \quad (67)$$

We take the boundary condition $x(z, t) \rightarrow x_{\text{good}}$ when $z \rightarrow -\infty$ and $x(z, t) \rightarrow x_{\text{bad}}$ when $z \rightarrow +\infty$. This boundary condition captures the profiles obtained after the transient phase has passed, and is well adapted to the study of the soliton propagation.

Velocity formula for scalar systems. As before, we assume that there exists a constant $v > 0$ such that, for $t \rightarrow +\infty$ and $|z - vt| = O(1)$, the profile $x(z, t) \rightarrow X(z - vt)$, where $X(z)$ is independent of the initial condition $x(z, 0)$ and satisfies $\lim_{z \rightarrow -\infty} X(z) = x_{\text{good}}$, $\lim_{z \rightarrow +\infty} X(z) = x_{\text{bad}}$. Under this assumption, the velocity of the soliton is

$$v = \frac{\Delta E}{\int_{\mathbb{R}} dz g'(X(z); \epsilon) X'(z)^2}, \quad (68)$$

where the shape $X(z)$ satisfies

$$X(z) - vX'(z) = \int_0^1 du f\left(\int_0^1 ds g(X(z-u+s); \epsilon); \epsilon\right) \quad (69)$$

C. Derivation of the velocity formula (68)

The derivation of (68) follows closely that in Section IV-B, so we will be quite brief. The first step is to introduce a continuum version of $U(\mathbf{x})$, which we call $\Delta\mathcal{U}(x(\cdot, \cdot))$. We define $x_0(z)$ as a static (time-independent) profile that satisfies the boundary conditions $x_0(z) \rightarrow x_{\text{good}}$ when $z \rightarrow -\infty$ and $x_0(z) \rightarrow x_{\text{bad}}$ when $z \rightarrow +\infty$ (for example, one may take a Heaviside step function). This is a reference profile in order to have well-defined integrals in the following expression

$$\begin{aligned} \Delta\mathcal{U}(x(\cdot, \cdot)) = \int_{\mathbb{R}} dz \Big[& \{x(z, t)g(x(z, t); \epsilon) - G(x(z, t); \epsilon) - F\left(\int_0^1 du g(x(z-u, t); \epsilon); \epsilon\right) \\ & - \{x_0(z)g(x_0(z); \epsilon) - G(x_0(z); \epsilon) - F\left(\int_0^1 du g(x_0(z-u); \epsilon); \epsilon\right)\} \Big]. \end{aligned}$$

As long as $x(z, t)$ and $x_0(z)$ converge to their limiting values fast enough, the integrals over the spatial axis are well defined. Evaluating the functional derivative⁴ of $\Delta\mathcal{U}[x(\cdot, \cdot); \epsilon]$ in an arbitrary direction $\eta(\cdot, \cdot)$, we find that (67) is equivalent to a gradient descent equation

$$\int_{\mathbb{R}} dz g'(x(z, t); \epsilon) (x(z, t+1) - x(z, t)) \eta(z, t) = -\frac{\delta\Delta\mathcal{U}}{\delta X}[\eta(z, t)] \quad (70)$$

Now we use the ansatz $x(z, t) \rightarrow X(z - vt)$ and apply (70) for the special direction $\eta(z, t) = X'(z - vt)$. Using also the approximation $X(z - vt) \approx X(z) - vX'(z)$ for small v we get (after a change of variables $z \rightarrow z + vt$)

$$v \int_{\mathbb{R}} dz X'(z)^2 g'(X(z); \epsilon) = \frac{\delta\Delta\mathcal{U}}{\delta x}[X'(z)]. \quad (71)$$

⁴Defined as $\lim_{\gamma \rightarrow 0} \gamma^{-1}(\Delta\mathcal{U}(x(\cdot, \cdot) + \gamma\eta(\cdot, \cdot)) - \Delta\mathcal{U}(x(\cdot, \cdot)))$

We then proceed to compute the right-hand side of (70). We split the potential functional into two parts: the “single system potential” $\mathcal{U}_s(x(\cdot, \cdot))$ that remains if we ignore the coupling effect, and the “interaction potential” $\mathcal{U}_i(x(\cdot, \cdot))$ that captures the effect of coupling. That is, $\Delta\mathcal{U} = \mathcal{U}_s + \mathcal{U}_i$, with

$$\begin{aligned}\mathcal{U}_s(x(\cdot, \cdot)) &= \int_{\mathbb{R}} dz \left[\{x(z, t)g(x(z, t); \epsilon) - G(x(z, t); \epsilon) - F(g(x(z, t); \epsilon))\right. \\ &\quad \left. - \{x_0(z)g(x_0(z); \epsilon) - G(x_0(z); \epsilon) - F(g(x_0(z); \epsilon); \epsilon)\} \right], \\ \mathcal{U}_i(x(\cdot, \cdot)) &= \int_{\mathbb{R}} dz \left[\{F(g(x(z, t); \epsilon); \epsilon) - F(\int_0^1 du g(x(z - u, t); \epsilon))\right. \\ &\quad \left. - \{F(g(x_0(z); \epsilon); \epsilon) - F(\int_0^1 du g(x_0(z - u); \epsilon); \epsilon)\} \right].\end{aligned}$$

The computation of each functional derivative at $x(z, t) \rightarrow X(z - vt)$ in the direction $X'(z - vt)$ yields

$$\begin{aligned}\frac{\delta \mathcal{U}_s}{\delta X}[X'(z)] &= \int_{\mathbb{R}} dz X'(z) \left(X(z)g'(X(z); \epsilon) - g'(X(z); \epsilon)f(g(X(z); \epsilon); \epsilon) \right) \\ &= \int_{\mathbb{R}} dz \frac{d}{dz} \left\{ X(z)g(X(z); \epsilon) - G(X(z); \epsilon) - F(g(X(z); \epsilon); \epsilon) \right\} \\ &= \left[U_s(X(z)) \right]_{-\infty}^{+\infty} \\ &= U_s(x_{\text{bad}}; \epsilon) - U_s(x_{\text{good}}; \epsilon).\end{aligned}\tag{72}$$

and

$$\begin{aligned}\frac{\delta \mathcal{U}_i}{\delta X}[X'(z)] &= \int_{\mathbb{R}} dz X'(z) \left\{ f(g(X(z); \epsilon); \epsilon)g'(X(z)) - f(\int_0^1 du g(X(z - u); \epsilon); \epsilon) \int_0^1 du g'(X(z - u); \epsilon) \right\} \\ &= \int_{\mathbb{R}} dz \frac{d}{dz} \left\{ F(g(X(z); \epsilon); \epsilon) - F(\int_0^1 du g(X(z - u); \epsilon); \epsilon) \right\} \\ &= 0.\end{aligned}\tag{73}$$

Replacing (72) and (73) in (70), we obtain the velocity formula (68).

D. Generalized LDPC (GLDPC) Codes

A GLDPC code is a code represented by a bipartite graph, such that the rules of the check nodes do not depend on parity (as do usual LDPC codes) but on a primitive BCH code. An attractive property of BCH codes is that they can be designed to correct a chosen number of errors. For instance, one can design a BCH code so that it corrects all patterns of at most e erasures on the BEC, and all error patterns of weight at most e on the binary symmetric channel (BSC). We consider a GLDPC code with degree-2 variable nodes and degree- n check nodes whose rules are given by a primitive BCH code of blocklength n .

We give a short description of a BCH code of blocklength n and minimum distance $d = 2e + 1$ (see [34] for more details). A BCH code is a cyclic code over a finite field $\text{GF}(b^\beta)$ where b is a prime power and β is an integer. Let a be a primitive element of $\text{GF}(b^\beta)$. Then each element of $\text{GF}(b^\beta)$ can be written in the form a^i , $i \in \mathbb{N}$. For each element a^i we can define a minimal polynomial $m_i(\cdot)$ which is the monic polynomial over $\text{GF}(b)$ with smallest degree. The generator polynomial $\theta(\cdot)$ over $\text{GF}(b)$ of the BCH code is defined as the least common multiple of $m_1(\cdot), \dots, m_d(\cdot)$.

Consider transmission on the BEC or BSC and denote by ϵ the channel parameter. The density evolution recursions have been derived in [35] for both channels, based on a bounded distance decoder for the BCH code. For n and e fixed, we can write the update equations of the message passing algorithm as (61) with

$$\begin{cases} f(x; \epsilon) = \epsilon x, \\ g(x) = \sum_{i=e}^{n-1} \binom{n-1}{i} x^i (1-x)^{n-i-1}. \end{cases}$$

Here, we have $\epsilon_{\max} = x_{\max} = y_{\max} = 1$. Moreover, one checks easily by differentiation (with respect to x) that the potential function $U_{\text{GLDPC}}(x)$ of the system is given by

$$U_{\text{GLDPC}}(x) = \frac{e}{n} g(x; \epsilon) - \frac{g'(x; \epsilon)}{n x (1-x)} - \frac{\epsilon}{2} g^2(x; \epsilon).$$

For numerical implementation purposes, it is useful to note that

$$g'(x) = \frac{x^{e-1} (1-x)^{n-e-1}}{B(e, n-e)},$$

where $B(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$ denotes the Beta function, and that $g(x)$ is equal to the regularized incomplete Euler Beta function so that

$$g(x) = \frac{1}{B(e, n-e)} \int_0^x ds s^{e-1} (1-s)^{n-e-1}.$$

This potential has $x = 0$ as a trivial stationary point (equivalently, this is a trivial fixed point of DE as can be seen from the expressions of f and g) and develops a non-trivial minimum at $x_{\text{BP}} \neq 0$ for $\epsilon > \epsilon_{\text{BP}}$. Note that ϵ_{BP} is found as usual as the first horizontal inflexion point. The MAP threshold is given by ϵ_{MAP} such that $U_{\text{GLDPC}}(x_{\text{BP}}) = U_{\text{GLDPC}}(0) = 0$.

The formula for the velocity of the soliton appearing for coupled GLDPC codes is found from (68). The energy gap for $\epsilon_{\text{BP}} \leq \epsilon \leq \epsilon_{\text{MAP}}$ is now $\Delta E = U_{\text{GLDPC}}(x_{\text{BP}}) - U_{\text{GLDPC}}(0)$. Figure 10 shows the velocities (normalized by w) for the spatially coupled GLPDC code with $n = 15$ and $e = 3$, when the coupling parameters satisfy $L_c + w = 500$ and $w = 3$. We plot the velocities for $\epsilon \in [\epsilon_{\text{BP}}, \epsilon_{\text{MAP}}] = [0.348, 0.394]$. We observe that the formula for the velocity provides a very good estimation of the empirical velocity v_e .

E. Compressive Sensing

Let \mathbf{s} be a length- n signal vector where the components are i.i.d. copies of a random variable S . We take m linear measurements of the signal and assume that the measurement matrix has i.i.d Gaussian elements distributed like $\mathcal{N}(0, 1/\sqrt{n})$. We define $\delta = m/n$ as the measurement ratio and fix it to a constant value when $n \rightarrow \infty$. The relation between δ and the parameter ϵ defined in Section VII-A is $\epsilon = \delta^{-1}$. We assume that the power of the variable S is normalized to 1; that is, $\mathbb{E}[S^2] = 1$. We also assume that each component of the signal \mathbf{s} is corrupted by independent Gaussian noise of variance $\sigma^2 = 1/\text{snr}$. To recover \mathbf{s} one implements the so-called approximate message passing (AMP) algorithm.

It is well-known that the analysis of the AMP algorithm is given by state evolution [10]. Let $Y = \sqrt{\text{snr}} S + Z$ where $Z \sim \mathcal{N}(0, 1)$, let $\hat{S}(Y) = \mathbb{E}_{S|Y}[S|Y]$ the minimum mean square estimator, and set

$$\text{mmse}(\text{snr}) = \mathbb{E}_{S,Y}[(S - \hat{S}(Y, \text{snr}))^2],$$

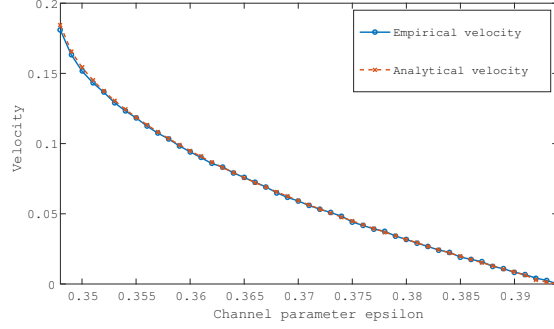


Fig. 10. We consider a GLDPC code with $n = 15$ and $e = 3$, with spatial length $L + w = 500$ and uniform coupling window with $w = 3$. We plot the normalized velocities v_{GLDPC} and v_e as a function of the channel parameter ϵ when ϵ is between the BP threshold $\epsilon_s = \epsilon_{\text{BP}} = 0.348$ and the potential threshold $\epsilon_c = \epsilon_{\text{MAP}} \approx 0.394$.

for the `mmse` function. The state evolution equations (which track the mean squared error of the AMP estimate) then correspond to the recursion (61) with

$$\begin{cases} f(x, \delta) = \text{mmse}(\text{snr} - x), \\ g(x, \delta) = \text{snr} - \frac{1}{\frac{1}{\text{snr}} + \frac{x}{\delta}}. \end{cases}$$

Here x is interpreted as the mean square error predicted by the AMP estimate of the signal. State evolution is initialized with $x = 1$ which corresponds to no knowledge about the signal. We will take δ as the control parameter. Note that we have $\delta_{\text{max}} = 1$, $x_{\text{max}} = \text{mmse}(0)$, $y_{\text{max}} = g(x_{\text{max}})$. The potential function is equal to

$$U_{\text{cs}}(x) = -\frac{x}{\frac{1}{\text{snr}} + \frac{x}{\delta}} + \delta \ln \left(1 + \frac{x \text{snr}}{\delta} \right) - 2I(S; \sqrt{\text{snr}}S + Z) + 2I\left(S; \sqrt{\frac{1}{\frac{1}{\text{snr}} + \frac{x}{\delta}}}S + Z\right),$$

where $I(A; B)$ is the mutual information between two random variables A and B . To check that this potential gives back the correct state evolution equation as a stationarity condition, we simply differentiate it with respect to x thanks to the well known relation $\frac{d}{d\text{snr}} I(S; \sqrt{\text{snr}}S + Z) = \frac{1}{2} \text{mmse}(\text{snr})$.

To illustrate the potential function in a concrete case we take the Bernoulli-Gaussian distribution as the prior distribution over the signal components

$$q_0(s) = (1 - \rho)\delta(s) + \rho \frac{e^{-s^2/2}}{\sqrt{2\pi}},$$

Figure 11 shows the potential function for $\rho = 0.1$, $\text{snr} = 10^5$, and several values of δ (the measurement fraction). We observe that, for $\delta > \delta_{\text{AMP}} = 0.208$, there is a unique minimum x_{good} which is a fixed point of state evolution when it is initialized to $x = 1$. In this phase, there are enough measurements so that the reconstruction of the signal is good and the mean square error is small. At $\delta_{\text{AMP}} = 0.208$ a horizontal inflexion point develops in the potential function. For $\delta < \delta_{\text{AMP}}$ a second minimum appears at a higher mean square error x_{bad} and the reconstruction of the AMP algorithm is bad. The optimal threshold corresponding to the minimum mean square error estimator is found when δ is such that the two minima of the potential are at the same height, namely δ_{opt} is given by the solution of the equation $U_{\text{sc}}(x_{\text{bad}}) = U_{\text{sc}}(x_{\text{good}})$. For our example one finds $\delta_{\text{opt}} = 0.157$. This threshold is reached by the AMP algorithm on the spatially coupled system.

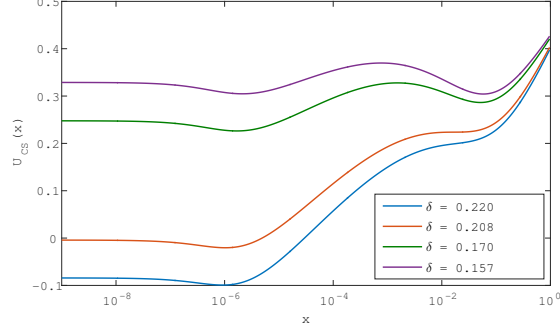


Fig. 11. Potential function for compressive sensing with a Gaussian-Bernoulli prior. The sparsity parameter is $\rho = 0.1$ and the signal to noise ratio $\text{snr} = 10^5$. We show the potential for several values of the measurement fraction δ . For $\delta > \delta_{\text{AMP}}$ we have a single minimum x_{good} . At $\delta_{\text{AMP}} = 0.208$ there is a horizontal inflexion point and for smaller measurement fractions a second minimum x_{bad} appears. At $\delta_{\text{opt}} = 0.157$ the gap ΔE vanishes.

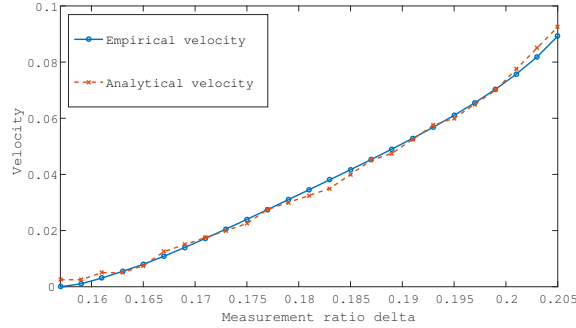


Fig. 12. We consider the compressive sensing problem with $\text{snr} = 10^5$ and Gaussian-Bernoulli prior for the signal components with sparsity parameter $\rho = 0.1$. We have $L_c + w = 250$ and uniform coupling window with $w = 4$. We plot the normalized velocities v_{CS} and v_e as a function of the measurement fraction δ when δ is between the potential threshold $\delta_{\text{opt}} = 0.157$ and $\delta_{\text{AMP}} = 0.208$.

Fix $\delta \in [\delta_{\text{opt}}, \delta_{\text{AMP}}] = [0.157, 0.208]$. In this regime spatially coupled state evolution develops a soliton which represents the profile of mean square error along the spatial direction. The formula for the velocity v_{CS} of this soliton is obtained from (68) where the energy gap is now $\Delta E = U_{\text{SC}}(x_{\text{bad}}) - U_{\text{SC}}(x_{\text{good}})$. Figure 12 shows the velocities (normalized by w) for the spatially coupled compressive sensing system with $\text{snr} = 10^5$, $\rho = 0.1$ and with the coupling parameters satisfy $L_c + w = 250$ and $w = 4$. We plot in Figure 12 the velocities for $\delta \in [\delta_{\text{opt}}, \delta_{\text{AMP}}] = [0.157, 0.208]$. It is clear that the formula for the velocity provides a good estimation of the empirical velocity v_e .

VIII. CONCLUSION

Our formulas for the velocities of the solitons that appear in spatially coupled codes for BMS channels and for general spatially coupled scalar systems (e.g. compressive sensing) rest on the approximation of the discrete system by a continuum one. We believe that this approximation becomes exact in an asymptotic limit of infinite spatial length and window size (keeping the order $L_c \gg w \gg 1$). It is an interesting open problem to quantify the quality

of this approximation already for L_c infinite and w finite but large. The numerical results tend to indicate that the approximation is already quite good for small values of w , when it is of the order of a few positions.

Another important and interesting open question concerns the proof of the ansatz, namely proving that the shape of the soliton is unique and independent of the initial condition on the profile. Settling this question would show that the velocity of the soliton is independent of the size of the seed that initiates decoding or signal reconstruction at the boundaries.

The formulas for the velocity involve the whole shape of the soliton in the denominator. For optimization purposes it would be desirable to have a good approximation (or bound) on the denominator that would involve only primitive quantities related to the underlying uncoupled ensemble (such as the degree distributions, the single system potential, etc). Such an approximation scheme has been proposed in [25] for the special case of the transmission over the BEC where it works quite well close to the MAP threshold. It would be desirable to find an extension to the more general situations considered in the present paper.

APPENDIX A

FUNCTIONAL DERIVATIVES

In this appendix we derive equations (21) and (32).

A. Derivation of Equ. (21)

We calculate the functional derivative of $\Delta\mathcal{W}(\mathbf{x})$ in a direction $\eta(z, t)$ as follows.

$$\frac{\delta\Delta\mathcal{W}}{\delta\mathbf{x}}[\eta(z, t)] = \frac{\partial}{\partial\gamma}\Delta\mathcal{W}(\mathbf{x} + \gamma\eta)\Big|_{\gamma=0} = \frac{\partial}{\partial\gamma}\int_{\mathbb{R}}dz P(z, \mathbf{x} + \gamma\eta)\Big|_{\gamma=0},$$

where the function $P(\cdot, \cdot)$ is defined in (19). Then, taking the derivative with respect to γ yields

$$\begin{aligned} \int_{\mathbb{R}}dz \Big\{ & H(\rho^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)) + H(\rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)) - H(\rho^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)) \\ & - H(\mathbf{x}(z, t) \boxtimes \rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)) - H\left(\mathbf{c}(z) \boxplus \lambda^{\boxplus}\left(\int_0^1 ds \rho^{\boxtimes}(\mathbf{x}(z + s, t))\right)\right. \\ & \left. \boxplus \left[\int_0^1 du \rho'^{\boxtimes}(\mathbf{x}(z + u, t)) \boxtimes \eta(z + u, t)\right]\right) \Big\}. \end{aligned}$$

We notice that the first and third terms in the integral cancel out due to the commutativity of the operator \boxtimes . By rearranging the averaging functions in the last term, we obtain

$$\begin{aligned} \int_{\mathbb{R}}dz \Big\{ & H(\rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)) - H(\mathbf{x}(z, t) \boxtimes \rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)) \\ & - H\left(\int_0^1 du \mathbf{c}(z - u) \boxplus \lambda^{\boxplus}\left(\int_0^1 ds \rho^{\boxtimes}(\mathbf{x}(z - u + s, t))\right) \boxplus \left[\rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)\right]\right) \Big\}. \end{aligned}$$

By noticing that $\mathbf{y} = \int_0^1 du \, c(z-u) \otimes \lambda^{\otimes} \left(\int_0^1 ds \, \rho^{\boxtimes}(\mathbf{x}(z-u+s, t)) \right)$ is a probability measure and $\mathbf{a} = \rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)$ is a difference of probability measures, we can use the second duality rule in (4) $H(\mathbf{y} \boxtimes \mathbf{a}) + H(\mathbf{y} \otimes \mathbf{a}) = H(\mathbf{a})$ to rewrite the above as, freely using the commutativity of \boxtimes ,

$$\begin{aligned} & \int_{\mathbb{R}} dz \left\{ -H(\mathbf{x}(z, t) \boxtimes \rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t)) + H \left(\left[\int_0^1 du \, c(z-u) \otimes \lambda^{\otimes} \left(\int_0^1 ds \, \rho^{\boxtimes}(\mathbf{x}(z-u+s, t)) \right) \right] \right. \right. \\ & \quad \left. \left. \boxtimes \rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t) \right) \right\} \\ &= \int_{\mathbb{R}} dz \, \rho'^{\boxtimes}(\mathbf{x}(z, t)) \boxtimes \eta(z, t) \boxtimes \left(\int_0^1 du \, c(z-u) \otimes \lambda^{\otimes} \left(\int_0^1 ds \, \rho^{\boxtimes}(\mathbf{x}(z-u+s, t)) \right) - \mathbf{x}(z, t) \right). \end{aligned}$$

B. Derivation of Equ. (32)

We calculate the functional derivative of $\mathcal{W}_s(\mathbf{X})$ in the direction $\mathbf{X}'(z)$ as follows.

$$\begin{aligned} \frac{\delta \mathcal{W}_s}{\delta \mathbf{X}}[\mathbf{X}'(z)] &= \frac{\partial}{\partial \gamma} \mathcal{W}_s(\mathbf{X} + \gamma \mathbf{X}') \Big|_{\gamma=0} = \frac{\partial}{\partial \gamma} \int_{\mathbb{R}} dz \, P_s(z, \mathbf{X} + \gamma \mathbf{X}') \Big|_{\gamma=0} \\ &= \int_{\mathbb{R}} dz \left\{ H(\rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)) + H(\rho'^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)) - H(\rho^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)) \right. \\ & \quad \left. - H(\mathbf{X}(z) \boxtimes \rho'^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)) - H(c(z) \otimes \lambda^{\otimes}(\rho^{\boxtimes}(\mathbf{X}(z))) \otimes [\rho'^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)]) \right\}. \end{aligned}$$

We notice here that on the right side of the last equality, the first and third terms under the integral cancel out. Using the second duality rule in (4), and noticing that $\mathbf{X}(z)$ is a probability measure and $\rho'^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)$ is a difference of probability measures, we can rewrite the functional derivative as

$$\int_{\mathbb{R}} dz \left\{ H(\mathbf{X}(z) \otimes [\rho'^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)]) - H(c(z) \otimes \lambda^{\otimes}(\rho^{\boxtimes}(\mathbf{X}(z))) \otimes [\rho'^{\boxtimes}(\mathbf{X}(z)) \boxtimes \mathbf{X}'(z)]) \right\}.$$

APPENDIX B

DERIVATION OF EXPRESSION (58)

Our goal in this appendix is to show that (57) reduces to (58) when $\delta \rightarrow 0$. We first reorganize (57) as follows (up to multiplication by $1/\delta^2$)

$$\begin{aligned} & \psi \left((r-2)\psi^{-1}(1-p(z)) + 2\psi^{-1}(1-p(z+\delta)) \right) - 2\psi \left((r-2)\psi^{-1}(1-p(z)) + \psi^{-1}(1-p(z+\delta)) \right. \\ & \quad \left. + \psi^{-1}(1-p(z)) \right) + \psi \left((r-2)\psi^{-1}(1-p(z)) + 2\psi^{-1}(1-p(z)) \right). \end{aligned}$$

When we Taylor expand each entropy $\psi(\dots)$ around

$$(r-2)\psi^{-1}(1-p(z))$$

to second order we observe that the first order terms cancel and what remains is

$$\begin{aligned} & -\frac{1}{2}\psi''((r-2)\psi^{-1}(1-p(z))) \left\{ 4\psi^{-1}(1-p(z+\delta))^2 - 4\psi^{-1}(1-p(z))^2 \right. \\ & \quad \left. - 2(\psi^{-1}(1-p(z+\delta)) - \psi^{-1}(1-p(z)))^2 \right\}. \end{aligned}$$

This is equal to

$$-\psi''((r-2)\psi^{-1}(1-p(z))) \times \left(\psi^{-1}(1-p(z+\delta)) - \psi^{-1}(1-p(z)) \right)^2.$$

Next, we write $\psi^{-1}(1 - p(z + \delta))$ as

$$\psi^{-1}(1 - p(z) - (p(z + \delta) - p(z)))$$

and Taylor expand $\psi^{-1}(\dots)$ around $1 - p(z)$ to obtain

$$-(p(z + \delta) - p(z))^2 ((\psi^{-1})'(1 - p(z)))^2 \times \psi''((r - 2)\psi^{-1}(1 - p(z))).$$

Multiplying by $1/\delta^2$, taking the limit $\delta \rightarrow 0$, and using the relation $(\psi^{-1})'(\dots) = 1/(\psi'(\psi^{-1}(\dots)))$ we obtain

$$-\left(\frac{dp(z)}{dz}\right)^2 \frac{\psi''((r - 2)\psi^{-1}(1 - p(z)))}{(\psi'(\psi^{-1}(1 - p(z))))^2}.$$

ACKNOWLEDGMENT

The authors would like to thank Ruediger Urbanke and Markus Stinner for discussions and suggestions on applications to finite-length scaling laws, and Tongxin Li for interactions on compressive sensing during his summer internship in EPFL.

REFERENCES

- [1] A. J. Felstrom and K. S. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Transactions on Information Theory*, pp. 2181–2190, 1999.
- [2] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7761–7813, 2013.
- [3] S. Kumar, A. J. Young, N. Macris, and H. D. Pfister, "Threshold saturation for spatially coupled LDPC and LDGM codes on BMS channels," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7389–7415, 2014.
- [4] M. Lentmaier, A. Sridharan, K. S. Zigangirov, and D. J. Costello Jr, "Terminated LDPC convolutional codes with thresholds close to capacity," in *International Symposium on Information Theory Proceedings (ISIT)*. IEEE, 2005, pp. 1372–1376.
- [5] M. Lentmaier, A. Sridharan, D. J. Costello Jr, and K. Zigangirov, "Iterative decoding threshold analysis for LDPC convolutional codes," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5274–5289, 2010.
- [6] V. Aref, N. Macris, R. Urbanke, and M. Vuffray, "Lossy source coding via spatially coupled LDGM ensembles," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 373–377.
- [7] V. Aref, N. Macris, and M. Vuffray, "Approaching the Rate-Distortion Limit With Spatial Coupling, Belief Propagation, and Decimation," *IEEE Transactions on Information Theory* 61 (7), 3954–3979, vol. 61, no. 7, pp. 3954–3979, 2015.
- [8] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2010, pp. 347–353.
- [9] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2012, no. 08, p. P08009, 2012.
- [10] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in *International Symposium on Information Theory Proceedings (ISIT)*. IEEE, 2012, pp. 1231–1235.
- [11] S. H. Hassani, N. Macris, and R. Urbanke, "Coupled graphical models and their thresholds," in *Proceedings of Information Theory Workshop IEEE (Dublin)*, 2010, pp. 1–5.
- [12] —, "Threshold saturation in spatially coupled constraint satisfaction problems," *Journal of Statistical Physics*, vol. 150, no. 5, pp. 807–850, 2013.
- [13] D. Achlioptas, S. H. Hassani, N. Macris, and R. Urbanke, "Bounds for random constraint satisfaction problems via spatial coupling," in *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, January 2016.
- [14] S. H. Hassani, N. Macris, and R. Urbanke, "Chains of mean field models," *Journal of Statistical Mechanics: Theory and Experiment*, vol. P02011, 2012.

- [15] F. Caltagirone, S. Franz, R. G. Morris, and L. Zdeborová, “Dynamics and termination cost of spatially coupled mean-field models,” *Physical Review E*, vol. 89, no. 1, p. 012102, 2014.
- [16] S. Kudekar, T. Richardson, and R. Urbanke, “Wave-like solutions of general one-dimensional spatially coupled systems,” *preprint arXiv:1208.5273*, 2012.
- [17] V. Aref, L. Schmalen, and S. ten Brink, “On the convergence speed of spatially coupled LDPC ensembles,” in *51st Annual Allerton Conference on Comm., Control, and Comp.* IEEE, 2013, pp. 342–349.
- [18] S.-Y. Chung and G. D. J. Forney, “On the capacity of low-density parity-check codes,” in *International Symposium on Information Theory Proceedings (ISIT)*. IEEE, 2001, p. 320.
- [19] S.-Y. Chung, “On the construction of some capacity-approaching coding schemes,” Ph.D. dissertation, MIT, 2000.
- [20] P. Olmos and R. Urbanke, “A scaling law to predict the finite-length performance of spatially-coupled LDPC codes,” *IEEE Transactions on Information Theory*, vol. 61, no. 6, pp. 3164–3184, 2014.
- [21] A. Yedla, Y.-Y. Jian, P. S. Nguyen, and H. D. Pfister, “A simple proof of Maxwell saturation for coupled scalar recursions,” *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6943–6965, 2014.
- [22] K. Takeuchi, T. Tanaka, and T. Kawabata, “A phenomenological study on threshold improvement via spatial coupling,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 95, no. 5, pp. 974–977, 2012.
- [23] R. El-Khatib, N. Macris, and R. Urbanke, “Displacement convexity, a useful framework for the study of spatially coupled codes,” in *Information Theory Workshop (ITW), Sevilla, Spain*. IEEE, 2013, pp. 1–5.
- [24] R. El-Khatib, N. Macris, T. Richardson, and R. Urbanke, “Analysis of coupled scalar systems by displacement convexity,” in *International Symposium on Information Theory Proceedings (ISIT), Honolulu, Hawaii*. IEEE, 2014, pp. 2321–2325.
- [25] R. El-Khatib and N. Macris, “The velocity of the decoding wave for spatially coupled codes on bms channels,” in *Information Theory Proceedings (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 2119–2123.
- [26] —, “The velocity of the propagating wave for general coupled scalar systems,” in *Information Theory Workshop, Cambridge, UK*, 2016.
- [27] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge University Press, 2008.
- [28] A. Giurgiu, N. Macris, and R. Urbanke, “Spatial coupling as a proof technique and three applications,” *IEEE Trans. Inform. Theory*, vol. 62, no. 10, pp. 5281–5295, 2016.
- [29] —, “How to prove the Maxwell conjecture via spatial coupling - A proof of concept,” in *International Symposium on Information Theory Proceedings (ISIT)*. IEEE, 2012, pp. 458–462.
- [30] —, “And now to something completely different: Spatial coupling as a proof technique,” in *International Symposium on Information Theory Proceedings (ISIT)*. IEEE, 2013, pp. 2443–2447.
- [31] S.-Y. Chung, T. Richardson, and R. Urbanke, “Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation,” *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 657–670, 2001.
- [32] J. Thorpe, “Low-density parity-check (LDPC) codes constructed from protographs,” *IPN progress report*, vol. 42, no. 154, pp. 42–154, 2003.
- [33] M. Stinner, L. Barletta, and P. M. Olmos, “Finite-length scaling based on belief propagation for spatially coupled LDPC codes,” *arXiv preprint arXiv:1604.05111*, 2016.
- [34] V. Pless, *Introduction to the theory of error-correcting codes*. John Wiley & Sons, 2011, vol. 48.
- [35] Y.-Y. Jian, H. D. Pfister, and K. R. Narayanan, “Approaching capacity at high rates with iterative hard-decision decoding,” in *International Symposium on Information Theory Proceedings (ISIT)*. IEEE, 2012, pp. 2696–2700.